RESEARCH CENTER

FIELD
Algorithmics, Programming, Software and Architecture

Activity Report 2019

Section New Results

Edition: 2020-03-21

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. Exchange algorithm for evaluation and approximation error-optimized polynomials

Machine implementation of mathematical functions often relies on polynomial approximations. The particularity is that rounding errors occur both when representing the polynomial coefficients on a finite number of bits, and when evaluating it in finite precision. Hence, for finding the best polynomial (for a given fixed degree, norm and interval), one has to consider both types of errors: approximation and evaluation. While efficient algorithms were already developed for taking into account the approximation error, the evaluation part is usually a posteriori handled, in an ad-hoc manner. In [15], we formulate a semi-infinite linear optimization problem whose solution is the best polynomial with respect to the supremum norm of the sum of both errors. This problem is then solved with an iterative exchange algorithm, which can be seen as an extension of the well-known Remez algorithm. A discussion and comparison of the obtained results on different examples are finally presented.

7.1.2. On Moment Problems with Holonomic Functions

Many reconstruction algorithms from moments of algebraic data were developed in optimization, analysis or statistics. Lasserre and Putinar proposed an exact reconstruction algorithm for the algebraic support of the Lebesgue measure, or of measures with density equal to the exponential of a known polynomial. Their approach relies on linear recurrences for the moments, obtained using Stokes theorem. In [16], we extend this study to measures with holonomic densities and support with real algebraic boundary. In the framework of holonomic distributions (i.e. they satisfy a holonomic system of linear partial or ordinary differential equations with polynomial coefficients), an alternate method to creative telescoping is proposed for computing linear recurrences for the moments. When the coefficients of a polynomial vanishing on the support boundary are given as parameters, the obtained recurrences have the advantage of staying linear with respect to them. This property allows for an efficient reconstruction method. Given a finite number of numerically computed moments for a measure with holonomic density, and assuming a real algebraic boundary for the support, we propose an algorithm for solving the inverse problem of obtaining both the coefficients of a polynomial vanishing on the boundary and those of the polynomials involved in the holonomic operators which annihilate the density.

7.1.3. A certificate-based approach to formally verified approximations

In [17], we present a library to verify rigorous approximations of univariate functions on real numbers, with the Coq proof assistant. Based on interval arithmetic, this library also implements a technique of validation a posteriori based on the Banach fixed-point theorem. We illustrate this technique on the case of operations of division and square root. This library features a collection of abstract structures that organize the specification of rigorous approximations, and modularize the related proofs. Finally, we provide an implementation of verified Chebyshev approximations, and we discuss a few examples of computations.

7.2. Floating-point and validated numerics

7.2.1. Error analysis of some operations involved in the Cooley-Tukey Fast Fourier Transform

We are interested in [4] in obtaining error bounds for the classical Cooley-Tukey 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.

7.2.2. Algorithms for triple-word arithmetic

Triple-word arithmetic consists in representing high-precision numbers as the unevaluated sum of three floating-point numbers (with “nonoverlapping” constraints that are explicited in the paper). We introduce and analyze in [7] various algorithms for manipulating triple-word numbers: rounding a triple-word number to a floating-point number, adding, multiplying, dividing, and computing square-roots of triple-word numbers, etc. We compare our algorithms, implemented in the Campary library, with other solutions of comparable accuracy. It turns out that our new algorithms are significantly faster than what one would obtain by just using the usual floating-point expansion algorithms in the special case of expansions of length 3.

7.2.3. Accurate Complex Multiplication in Floating-Point Arithmetic

We deal in [24] with accurate complex multiplication in binary floating-point arithmetic, with an emphasis on the case where one of the operands in a “double-word” number. We provide an algorithm that returns a complex product with normwise relative error bound close to the best possible one, i.e., the rounding unit u.

7.2.4. Semi-automatic implementation of the complementary error function

The normal and complementary error functions are ubiquitous special functions for any mathematical library. They have a wide range of applications. Practical applications call for customized implementations that have strict accuracy requirements. Accurate numerical implementation of these functions is, however, non-trivial. In particular, the complementary error function erfc for large positive arguments heavily suffers from cancellation, which is largely due to its asymptotic behavior. We provide a semi-automatic code generator for the erfc function which is parameterized by the user-given bound on the relative error. Our solution, presented in [31], exploits the asymptotic expression of erfc and leverages the automatic code generator Metalibm that provides accurate polynomial approximations. A fine-grained a priori error analysis provides a libm developer with the required accuracy for each step of the evaluation. In critical parts, we exploit double-word arithmetic to achieve implementations that are fast, yet accurate up to 50 bits, even for large input arguments. We demonstrate that for high required accuracies the automatically generated code has performance comparable to that of the standard libm and for lower ones our code demonstrated roughly 25% speedup.

7.2.5. Posits: the good, the bad and the ugly

Many properties of the IEEE-754 floating-point number system are taken for granted in modern computers and are deeply embedded in compilers and low-level softare routines such as elementary functions or BLAS. In [32] we review such properties on the recently proposed Posit number system. Some are still true. Some are no longer true, but sensible work-arounds are possible, and even represent exciting challenge for the community. Some, in particular the loss of scale invariance for accuracy, are extremely dangerous if Posits are to replace floating point completely. This study helps framing where Posits are better than floating-point, where they are worse, and what tools are missing in the Posit landscape. For general-purpose computing, using Posits as a storage format only could be a way to reap their benefits without loosing those of classical floating-point. The hardware cost of this alternative is studied.

7.2.6. The relative accuracy of $(x+y)*(x-y)$

We consider in [8] the relative accuracy of evaluating $(x+y)(x-y)$ in IEEE floating-point arithmetic, when x and y are two floating-point numbers and rounding is to nearest. This expression can be used for example as an efficient cancellation-free alternative to x^2-y^2 and is well known to have low relative error, namely, at most about $3u$ with u denoting the unit roundoff. In this paper we complement this traditional analysis with a finer-grained one, aimed at improving and assessing the quality of that bound. Specifically, we show that if the tie-breaking rule is *to away* then the bound $3u$ is asymptotically optimal. In contrast, if the tie-breaking rule is *to even*, we show that asymptotically optimal bounds are now $2.25u$ for base two and $2u$ for larger bases, such as base ten. In each case, asymptotic optimality is obtained by the explicit construction of a certificate, that is, some floating-point input (x,y) parametrized by u and such that the error of the associated result is equivalent to the error bound as $u\to 0$. We conclude with comments on how $(x+y)(x-y)$ compares with x^2 in the presence of floating-point arithmetic, in particular showing cases where the computed value of $(x+y)(x-y)$ exceeds that of x^2.

7.2.7. The MPFI Library: Towards IEEE 1788-2015 Compliance

The IEEE 1788-2015 has standardized interval arithmetic. However, few libraries for interval arithmetic are compliant with this standard. In the first part of [30], the main features of the IEEE 1788-2015 standard are detailed. These features were not present in the libraries developed prior to the elaboration of the standard. MPFI is such a library: it is a C library, based on MPFR, for arbitrary precision interval arithmetic. MPFI is not (yet) compliant with the IEEE 1788-2015 standard for interval arithmetic: the planned modifications are presented.

7.3. Lattices: algorithms and cryptology

7.3.1. Approx-SVP in ideal lattices with pre-processing

In [28], we describe an algorithm to solve the approximate Shortest Vector Problem for lattices corresponding to ideals of the ring of integers of an arbitrary number field K. This algorithm has a pre-processing phase, whose run-time is exponential in $\log |\Delta|$ with Δ the discriminant of K. Importantly, this pre-processing phase depends only on K. The pre-processing phase outputs an advice, whose bit-size is no more than the run-time of the query phase. Given this advice, the query phase of the algorithm takes as input any ideal I of the ring of integers, and outputs an element of I which is at most $\exp(\widetilde{O}((\log |\Delta|)^{\alpha+1}/n))$ times longer than a shortest non-zero element of I (with respect to the Euclidean norm of its canonical embedding). This query phase runs in time and space $\exp(\widetilde{O}((\log |\Delta|)^{\max(2/3, 1-2\alpha)}))$ in the classical setting, and $\exp(\widetilde{O}((\log |\Delta|)^{1-2\alpha}))$ in the quantum setting. The parameter α can be chosen arbitrarily in $[0, 1/2]$. Both correctness and cost analyses rely on heuristic assumptions, whose validity is consistent with experiments.

The algorithm builds upon the algorithms from Cramer al. [EUROCRYPT 2016] and Cramer et al. [EUROCRYPT 2017]. It relies on the framework from Buchmann [Séminaire de théorie des nombres 1990], which allows to merge them and to extend their applicability from prime-power cyclotomic fields to all number fields. The cost improvements are obtained by allowing precomputations that depend on the field only.

7.3.2. An LLL algorithm for module lattices

The LLL algorithm takes as input a basis of a Euclidean lattice, and, within a polynomial number of operations, it outputs another basis of the same lattice but consisting of rather short vectors. In [23], we provide a generalization to R-modules contained in K^n for arbitrary number fields K and dimension n, with R denoting the ring of integers of K. Concretely, we introduce an algorithm that efficiently finds short vectors in rank-n modules when given access to an oracle that finds short vectors in rank-2 modules, and an algorithm that efficiently finds short vectors in rank-2 modules given access to a Closest Vector Problem oracle for a lattice that depends only on K. The second algorithm relies on quantum computations and its analysis is heuristic.

7.3.3. The general sieve kernel and new records in lattice reduction

In [14], we propose the General Sieve Kernel (G6K), an abstract stateful machine supporting a wide variety of lattice reduction strategies based on sieving algorithms. Using the basic instruction set of this abstract stateful machine, we first give concise formulations of previous sieving strategies from the literature and then propose new ones. We then also give a light variant of BKZ exploiting the features of our abstract stateful machine. This encapsulates several recent suggestions (Ducas at Eurocrypt 2018; Laarhoven and Mariano at PQCrypto 2018) to move beyond treating sieving as a blackbox SVP oracle and to utilise strong lattice reduction as preprocessing for sieving. Furthermore, we propose new tricks to minimise the sieving computation required for a given reduction quality with mechanisms such as recycling vectors between sieves, on-the-fly lifting and flexible insertions akin to Deep LLL and recent variants of Random Sampling Reduction.

Moreover, we provide a highly optimised, multi-threaded and tweakable implementation of this machine which we make open-source. We then illustrate the performance of this implementation of our sieving strategies by applying G6K to various lattice challenges. In particular, our approach allows us to solve previously unsolved instances of the Darmstadt SVP (151, 153, 155) and LWE (e.g. (75, 0.005)) challenges. Our solution for the SVP-151 challenge was found 400 times faster than the time reported for the SVP-150 challenge, the previous record. For exact SVP, we observe a performance crossover between G6K and FPLLL's state of the art implementation of enumeration at dimension 70.

7.3.4. Statistical zeroizing attack: cryptanalysis of candidates of BP obfuscation over GGH15 multilinear map

In [19], we present a new cryptanalytic algorithm on obfuscations based on GGH15 multilinear map. Our algorithm, statistical zeroizing attack, directly distinguishes two distributions from obfuscation while it follows the zeroizing attack paradigm, that is, it uses evaluations of zeros of obfuscated programs.

Our attack breaks the recent indistinguishability obfuscation candidate suggested by Chen et al. (CRYPTO'18) for the optimal parameter settings. More precisely, we show that there are two functionally equivalent branching programs whose CVW obfuscations can be efficiently distinguished by computing the sample variance of evaluations.

This statistical attack gives a new perspective on the security of the indistinguishability obfuscations: we should consider the shape of the distributions of evaluation of obfuscation to ensure security.

In other words, while most of the previous (weak) security proofs have been studied with respect to algebraic attack model or ideal model, our attack shows that this algebraic security is not enough to achieve indistinguishability obfuscation. In particular, we show that the obfuscation scheme suggested by Bartusek et al. (TCC'18) does not achieve the desired security in a certain parameter regime, in which their algebraic security proof still holds.

The correctness of statistical zeroizing attacks holds under a mild assumption on the preimage sampling algorithm with a lattice trapdoor. We experimentally verify this assumption for implemented obfuscation by Halevi et al. (ACM CCS'17).

7.3.5. Cryptanalysis of the CLT13 multilinear map

The reference [6] is the journal version of the Eurocrypt'15 article with the same title and authors.

7.3.6. Multi-Client Functional Encryption for Linear Functions in the Standard Model from LWE

Multi-client functional encryption (MCFE) allows ℓ clients to encrypt ciphertexts $\mathbf{C}_{t,1}, \mathbf{C}_{t,2}, ..., \mathbf{C}_{t,\ell}$ under some label. Each client can encrypt his own data X_i for a label t using a private encryption key ek_i issued by a trusted authority in such a way that, as long as all $\mathbf{C}_{t,i}$ share the same label t, an evaluator endowed with a functional key dk_f can evaluate $f(X_1, X_2, ..., X_\ell)$ without learning anything else on the underlying plaintexts X_i. Functional decryption keys can be derived by the central authority using the master secret key. Under the Decision Diffie-Hellman assumption, Chotard *et al.* (Asiacrypt 2018) recently described an adaptively secure MCFE scheme for the evaluation of linear functions over the integers. They also gave a decentralized variant (DMCFE) of their scheme which does not rely on a centralized authority, but rather allows encryptors to issue functional secret keys in a distributed manner. While efficient, their constructions both rely on random oracles in their security analysis. In [27], we build a standard-model MCFE scheme for the same functionality and prove it fully secure under adaptive corruptions. Our proof relies on the Learning-With-Errors (LWE) assumption and does not require the random oracle model. We also provide a decentralized variant of our scheme, which we prove secure in the static corruption setting (but for adaptively chosen messages) under the LWE assumption.

7.3.7. Zero-Knowledge Elementary Databases with More Expressive Queries

Zero-knowledge elementary databases (ZK-EDBs) are cryptographic schemes that allow a prover to commit to a set D of key-value pairs so as to be able to prove statements such as “x belongs to the support of D and D(x) = y” or “x is not in the support of D”. Importantly , proofs should leak no information beyond the proven statement and even the size of D should remain private. Chase et al. (Eurocrypt'05) showed that ZK-EDBs are implied by a special flavor of non-interactive commitment, called mercurial commitment, which enables efficient instantiations based on standard number theoretic assumptions. On the other hand, the resulting ZK-EDBs are only known to support proofs for simple statements like (non-)membership and value assignments. In [25], we show that mercurial commitments actually enable significantly richer queries. We show that, modulo an additional security property met by all known efficient constructions, they actually enable range queries over keys and values-even for ranges of super-polynomial size-as well as membership/non-membership queries over the space of values. Beyond that, we exploit the range queries to realize richer queries such as k-nearest neighbors and revealing the k smallest or largest records within a given range. In addition, we provide a new realization of trapdoor mercurial commitment from standard lattice asssumptions, thus obtaining the most expressive quantum-safe ZK-EDB construction so far.

7.3.8. Lossy Algebraic Filters With Short Tags

Lossy algebraic filters (LAFs) are function families where each function is parametrized by a tag, which determines if the function is injective or lossy. While initially introduced by Hofheinz (Eurocrypt 2013) as a technical tool to build encryption schemes with key-dependent message chosen-ciphertext (KDM-CCA) security, they also find applications in the design of robustly reusable fuzzy extractors. So far, the only known LAF family requires tags comprised of $\Theta(n^2)$ group elements for functions with input space $\mathbb{Z}_p$, where p is the group order. In [26], we describe a new LAF family where the tag size is only linear in n and prove it secure under simple assumptions in asymmetric bilinear groups. Our construction can be used as a drop-in replacement in all applications of the initial LAF system. In particular, it can shorten the ciphertexts of Hofheinz's KDM-CCA-secure public-key encryption scheme by 19 group elements. It also allows substantial space improvements in a recent fuzzy extractor proposed by Wen and Liu (Asiacrypt 2018). As a second contribution , we show how to modify our scheme so as to prove it (almost) tightly secure, meaning that security reductions are not affected by a concrete security loss proportional to the number of adversarial queries.

7.3.9. Shorter Quadratic QA-NIZK Proofs

Despite recent advances in the area of pairing-friendly Non-Interactive Zero-Knowledge proofs, there have not been many efficiency improvements in constructing arguments of satisfiability of quadratic (and larger degree) equations since the publication of the Groth-Sahai proof system (J. of Cryptology 2012). In [20], we address the problem of aggregating such proofs using techniques derived from the interactive setting and recent constructions of SNARKs. For certain types of quadratic equations, this problem was investigated before by González et al. (Asiacrypt'15). Compared to their result, we reduce the proof size by approximately 50

7.3.10. Shorter Pairing-based Arguments under Standard Assumptions

The paper [22] constructs efficient non-interactive arguments for correct evaluation of arithmetic and Boolean circuits with proof size $O(d)$ group elements, where d is the multiplicative depth of the circuit, under falsifiable assumptions. This is achieved by combining techniques from SNARKs and QA-NIZK arguments of membership in linear spaces. The first construction is very efficient (the proof size is $\approx 4d$ group elements and the verification cost is $4d$ pairings and $O(n+n+d)$ exponentiations, where n is the size of the input and n of the output) but one type of attack can only be ruled out assuming the knowledge soundness of QA-NIZK arguments of membership in linear spaces. We give an alternative construction which replaces this assumption with a decisional assumption in bilinear groups at the cost of approximately doubling the proof size. The construction for Boolean circuits can be made zero-knowledge with Groth-Sahai proofs, resulting in a NIZK argument for circuit satisfiability based on falsifiable assumptions in bilinear groups of proof size $O(n+d)$. Our main technical tool is what we call an "argument of knowledge transfer". Given a commitment C_1 and an opening x, such an argument allows to prove that some other commitment C_2 opens to $f(x)$, for some function f, even if C_2 is not extractable. We construct very short, constant-size, pairing-based arguments of knowledge transfer with constant-time verification for any linear function and also for Hadamard products. These allow to transfer the knowledge of the input to lower levels of the circuit.

7.3.11. Shorter Ring Signatures from Standard Assumptions

Ring signatures, introduced by Rivest, Shamir and Tauman (ASIACRYPT 2001), allow to sign a message on behalf of a set of users while guaranteeing authenticity and anonymity. Groth and Kohlweiss (EUROCRYPT

2015) and Libert *et al.* (EUROCRYPT 2016) constructed schemes with signatures of size logarithmic in the number of users. An even shorter ring signature, of size independent from the number of users, was recently proposed by Malavolta and Schroeder (ASIACRYPT 2017). However, all these short signatures are obtained relying on strong and controversial assumptions. Namely, the former schemes are both proven secure in the random oracle model while the later requires non-falsifiable assumptions.

The most efficient construction under mild assumptions remains the construction of Chandran et al. (ICALP 2007) with a signature of size $\Theta(\sqrt{n})$, where n is the number of users, and security is based on the Diffie-Hellman assumption in bilinear groups (the SXDH assumption in asymmetric bilinear groups).

In [21], we construct an asymptotically shorter ring signature from the hardness of the Diffie-Hellman assumption in bilinear groups. Each signature comprises $\Theta(n^{1/3})$ group elements, signing a message requires computing $\Theta(n^{1/3})$ exponentiations, and verifying a signature requires $\Theta(n^{2/3})$ pairing operations.

7.3.12. Two-Party ECDSA from Hash Proof Systems and Efficient Instantiations

ECDSA is a widely adopted digital signature standard. Unfortunately, efficient distributed variants of this primitive are notoriously hard to achieve and known solutions often require expensive zero knowledge proofs to deal with malicious adversaries. For the two party case, Lindell (CRYPTO 2017) recently managed to get an efficient solution which, to achieve simulation-based security, relies on an interactive, non standard, assumption on Paillier's cryptosystem.

In this paper [18] we generalize Lindell's solution using hash proof systems. The main advantage of our generic method is that it results in a simulation-based security proof without resorting to non-standard interactive assumptions.

Moving to concrete constructions, we show how to instantiate our framework using class groups of imaginary quadratic fields. Our implementations show that the practical impact of dropping such interactive assumptions is minimal. Indeed, while for 128-bit security our scheme is marginally slower than Lindell's, for 256-bit security it turns out to be better both in key generation and signing time. Moreover, in terms of communication cost, our implementation significantly reduces both the number of rounds and the transmitted bits without exception.

7.3.13. Algebraic XOR-RKA-Secure Pseudorandom Functions from Post-Zeroizing Multilinear Maps

In [13], we construct the first pseudorandom functions that resist a strong class of attacks where an adversary is able to run the cryptosystem not only with the fixed secret key, but with related keys where bits of its choice of the original keys are flipped. This problem is motivated by practical attacks that have been performed against physical devices. Our construction guarantees that every output of our construction, for the original key or for tampered keys, are pseudorandom, i.e. are computationally hard to distinguish from truly random values. To achieve this, we rely on a recent tool introduced in cryptography and termed multilinear maps. While multilinear maps have been recently attacked by several techniques, we prove that our construction remains secure despite the numerous vulnerabilities of current constructions of multilinear maps.

7.3.14. Unifying Leakage Models on a Rényi Day

Most theoretical models in cryptography suppose that an attacker can only observe the input/output behavior of a cryptosystem and nothing more. Yet, in the real world, cryptosystems run on physical devices and auxiliary information leaks from these devices. This leakage can sometimes be used to attack the system, even though it is proven secure in theory. To circumvent these issues, cryptographers have introduces several new security models in an attempt to encompass the different forms of leakage. Some models are simple, such as the probing model, and simple compilers allow to transform a system into one secure in the probing model, while some more realistic problems such as the noisy-leakage model are very involved. In [29], we show that these models are actually equivalent, proving in particular that the simple compilers are sufficient to guarantee security in realistic environments.

7.4. Algebraic computing and high-performance kernels

7.4.1. 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].

7.4.2. Absolute root separation

The absolute separation of a polynomial is the minimum nonzero difference between the absolute values of its roots. In the case of polynomials with integer coefficients, it can be bounded from below in terms of the degree and the height (the maximum absolute value of the coefficients) of the polynomial. We improve the known bounds for this problem and related ones. Then we report on extensive experiments in low degrees, suggesting that the current bounds are still very pessimistic. [5]

7.4.3. Improving the complexity of block low-rank factorizations with fast matrix arithmetic

We consider in [9] the LU factorization of an $n \times n$ matrix represented as a block low-rank (BLR) matrix: most of its off-diagonal blocks are approximated by matrices of small rank r, which reduces the asymptotic complexity of computing the LU factorization down to $\mathcal{O}(n^2 r)$. Even though lower complexities can be achieved with hierarchical matrices, the BLR format allows for a very simple and efficient implementation. In this article, our aim is to further reduce the BLR complexity without losing its nonhierarchical nature by exploiting fast matrix arithmetic, that is, the ability to multiply two $n \times n$ full-rank matrices together for $\mathcal{O}(n^{\omega})$ flops, where $\omega < 3$. We devise a new BLR factorization algorithm whose cost is $\mathcal{O}(n^{(\omega+1)/2} r^{(\omega-1)/2})$, which represents an asymptotic improvement compared with the standard BLR factorization as soon as $\omega < 3$. In particular, for Strassen’s algorithm, $\omega \approx 2.81$ yields the cost $\mathcal{O}(n^{1.904} r^{0.904})$. Our numerical experiments are in good agreement with this analysis.

7.4.4. Fast computation of approximant bases in canonical form

In [10] 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.

AROMATH Project-Team

5. New Results

5.1. Truncated Normal Forms for Solving Polynomial Systems: Generalized and Efficient Algorithms

Participant: Bernard Mourrain.

In [16], 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 $\mathbb{C}$. Normal form algorithms provide an algebraic approach to solve this problem. The framework presented in Telen et al. (2018) uses truncated normal forms (TNFs) to compute the algebra structure of R/I and the solutions of I. This framework allows for the use of much more general bases than the standard monomials for R/I. This is exploited in this paper to introduce the use of two special (non-monomial) types of basis functions with nice properties. This allows, for instance, to adapt the basis functions to the expected location of the roots of I. We also propose algorithms for efficient computation of TNFs and a generalization of the construction of TNFs in the case of non-generic zero-dimensional systems. The potential of the TNF method and usefulness of the new results are exposed by many experiments.

This is a joint work with Simon Telen and Marc Van Barel, Department of Computer Science - K.U.Leuven.

5.2. Implicit representations of high-codimension varieties

Participants: Ioannis Emiris, Clément Laroche, Christos Konaxis.

In [8], we study implicitization, which usually focuses on plane curves and (hyper)surfaces, in other words, varieties of codimension 1. We shift the focus on space curves and, more generally, on varieties of codimension larger than 1, and discuss approaches that are not sensitive to base points. Our first contribution is a direct generalization of an implicitization method based on interpolation matrices for objects of high codimension given parametrically or as point clouds. Our result shows the completeness of this approach which, furthermore, reduces geometric operations and predicates to linear algebra computations. Our second, and main contribution is an implicitization method of parametric space curves and varieties of codimension > 1, which exploits the theory of Chow forms to obtain the equations of conical (hyper)surfaces intersecting precisely at the given object. We design a new, practical, randomized algorithm that always produces correct output but possibly with a non-minimal number of surfaces. For space curves, which is the most common case, our algorithm returns 3 surfaces whose polynomials are of near-optimal degree; moreover, computation reduces to a Sylvester resultant. We illustrate our algorithm through a series of examples and compare our Maple code with other methods implemented in Maple. Our prototype is not faster but yields fewer equations and is more robust than Maple's implicitize. Although not optimized, it is comparable with Gröbner bases and matrix representations derived from syzygies, for degrees up to 6.

5.3. Saturation of Jacobian ideals: Some applications to nearly free curves, line arrangements and rational cuspidal plane curves

Participant: Alexandru Dimca.

In [6] we describe the minimal resolution of the ideal I_f, the saturation of the Jacobian ideal of a nearly free plane curve $(C:f)=0$. In particular, it follows that this ideal I_f can be generated by at most 4 polynomials. Some applications to rational cuspidal plane curves are given, and a natural related question is raised.

This is a joint work with Gabriel Sticlaru (Ovidius University of Constanta).

5.4. Matrix formulae for Resultants and Discriminants of Bivariate Tensor-product Polynomials

Participants: Laurent Busé, Angelos Mantzaflaris.

The construction of optimal resultant formulae for polynomial systems is one of the main areas of research in computational algebraic geometry. However, most of the constructions are restricted to formulae for unmixed polynomial systems, that is, systems of polynomials which all have the same support. Such a condition is restrictive, since mixed systems of equations arise frequently in many problems. Nevertheless, resultant formulae for mixed polynomial systems is a very challenging problem. In [5], we introduce a square, Koszul-type, matrix, the determinant of which is the resultant of an arbitrary (mixed) bivariate tensor-product polynomial system. The formula generalizes the classical Sylvester matrix of two univariate polynomials, since it expresses a map of degree one, that is, the elements of the corresponding matrix are up to sign the coefficients of the input polynomials. Interestingly, the matrix expresses a primal-dual multiplication map, that is, the tensor product of a univariate multiplication map with a map expressing derivation in a dual space. In addition we prove an impossibility result which states that for tensor-product systems with more than two (affine) variables there are no universal degree-one formulae, unless the system is unmixed. Last but not least, we present applications of the new construction in the efficient computation of discriminants and mixed discriminants.

This is joint work with Elias Tsigaridas (Ouragan, Inria).

5.5. Implicitizing rational curves by the method of moving quadrics

Participants: Laurent Busé, Clément Laroche, Fatmanur Yildirim.

In [4], a new technique for finding implicit matrix-based representations of rational curves in arbitrary dimension is introduced. It relies on the use of moving quadrics following curve parameterizations, providing a high-order extension of the implicit matrix representations built from their linear counterparts, the moving planes. The matrices we obtain offer new, more compact, implicit representations of rational curves. Their entries are filled by linear and quadratic forms in the space variables and their ranks drop exactly on the curve. Typically, for a general rational curve of degree d we obtain a matrix whose size is half of the size of the corresponding matrix obtained with the moving planes method. We illustrate the advantages of these new matrices with some examples, including the computation of the singularities of a rational curve.

5.6. Separation bounds for polynomial systems

Participants: Ioannis Emiris, Bernard Mourrain.

In [9] we rely on aggregate separation bounds for univariate polynomials to introduce novel worst-case separation bounds for the isolated roots of zero-dimensional, positive-dimensional, and overdetermined polynomial systems. We exploit the structure of the given system, as well as bounds on the height of the sparse (or toric) resultant, by means of mixed volume, thus establishing adaptive bounds. Our bounds improve upon Canny's Gap theorem [9]. Moreover, they exploit sparseness and they apply without any assumptions on the input polynomial system. To evaluate the quality of the bounds, we present polynomial systems whose root separation is asymptotically not far from our bounds. We apply our bounds to three problems. First, we use them to estimate the bit-size of the eigenvalues and eigenvectors of an integer matrix; thus we provide a new proof that the problem has polynomial bit complexity. Second, we bound the value of a positive polynomial over the simplex: we improve by at least one order of magnitude upon all existing bounds. Finally, we asymptotically bound the number of steps of any purely subdivision-based algorithm that isolates all real roots of a polynomial system.

This is a joint work with E. Tsigaridas (Ouragan).

5.7. Sparse polynomial interpolation: sparse recovery, super resolution, or Prony?

Participant: Bernard Mourrain.

In [12], we show that the sparse polynomial interpolation problem reduces to a discrete super-resolution problem on the n-dimensional torus. Therefore the semidefinite programming approach initiated by Candès & Fernandez-Granda in the univariate case can be applied. We extend their result to the multivariate case, i.e., we show that exact recovery is guaranteed provided that a geometric spacing condition on the supports holds and the number of evaluations are sufficiently many (but not many). It also turns out that the sparse recovery LP-formulation of $\ell 1$-norm minimization is also guaranteed to provide exact recovery *provided that* the evaluations are made in a certain manner and even though the Restricted Isometry Property for exact recovery is not satisfied. (A naive sparse recovery LP-approach does not offer such a guarantee.) Finally we also describe the algebraic Prony method for sparse interpolation, which also recovers the exact decomposition but from less point evaluations and with no geometric spacing condition. We provide two sets of numerical experiments, one in which the super-resolution technique and Prony's method seem to cope equally well with noise, and another in which the super-resolution technique seems to cope with noise better than Prony's method, at the cost of an extra computational burden (i.e. a semidefinite optimization).

This is a joint work with Cédric Josz and Jean-Bernard Lasserre (Équipe Méthodes et Algorithmes en Commande, LAAS).

5.8. Computing minimal Gorenstein covers

Participant: Bernard Mourrain.

In [7], we analyze and present an effective solution to the minimal Gorenstein cover problem: given a local Artin k–algebra $A = k[[x_1, ..., x_n]]/I$, compute an Artin Gorenstein k–algebra $G = k[[x_1, ..., x_n]]/I$ such that $\ell(G) - \ell(A)$ is minimal. We approach the problem by using Macaulay's inverse systems and a modification of the integration method for inverse systems to compute Gorenstein covers. We propose new characterizations of the minimal Gorenstein cover and present a new algorithm for the effective computation of the variety of all minimal Gorenstein covers of A for low Gorenstein colength. Experimentation illustrates the practical behavior of the method.

This is a joint work with Juan Elias and Roser Homs (Dep. de Matematiques i Informatica, Universitat de Barcelona).

5.9. Symmetry Preserving Interpolation

Participants: Erick David Rodriguez Bazan, Evelyne Hubert.

In [22], we address multivariate interpolation in the presence of symmetry. Interpolation is a prime tool in algebraic computation while symmetry is a qualitative feature that can be more relevant to a mathematical model than the numerical accuracy of the parameters. The article shows how to exactly preserve symmetry in multivariate interpolation while exploiting it to alleviate the computational cost. We revisit minimal degree and least interpolation with symmetry adapted bases, rather than monomial bases.This allows to construct bases of invariant interpolation spaces in blocks, capturing the inherent redundancy in the computations.We show that the so constructed symmetry adapted interpolation bases alleviate the computational cost of any interpolation problem and automatically preserve any equivariance of their interpolation problem might have.

5.10. Skew-Symmetric Tensor Decomposition

Participant: Bernard Mourrain.

In [2], we introduce the "skew apolarity lemma" and we use it to give algorithms for the skew-symmetric rank and the decomposition of tensors in $\wedge^d V_{\mathbb{C}}$ with $d \leq 3$ and $\dim V_{\mathbb{C}} \leq 8$. New algorithms to compute the rank and a minimal decomposition of a tri-tensor are also presented.

This is a joint work with Enrique Arrondo (UCM - Universidad Complutense de Madrid, Spain), Alessandra Bernardi (Department of Mathematics, University of Trento, Italy) Pedro Macias Marques (Departamento de Matemática da Universidade de Évora, Spain).

5.11. On the maximal number of real embeddings of minimally rigid graphs in $\mathbb{R}^2$, $\mathbb{R}^3$ and $\mathbb{S}^2$

Participants: Ioannis Emiris, Evangelos Bartzos.

In [3], we study the Rigidity theory studies the properties of graphs that can have rigid embeddings in the d-dimensional Euclidean space, or on a sphere and other manifolds which in addition satisfy certain edge length constraints. One of the major open problems in this field is to determine lower and upper bounds on the number of realizations with respect to a given number of vertices. This problem is closely related to the classification of rigid graphs according to their maximal number of real embeddings. In this paper, we are interested in finding edge lengths that can maximize the number of real embeddings of minimally rigid graphs in the plane, space, and on the sphere. We use algebraic formulations to provide upper bounds. To find values of the parameters that lead to graphs with a large number of real realizations, possibly attaining the (algebraic) upper bounds, we use some standard heuristics and we also develop a new method inspired by coupler curves. We apply this new method to obtain embeddings in $\mathbb{R}^3$. One of its main novelties is that it allows us to sample efficiently from a larger number of parameters by selecting only a subset of them at each iteration. Our results include a full classification of the 7-vertex graphs according to their maximal numbers of real embeddings in the cases of the embeddings in $\mathbb{R}^2$ and $\mathbb{R}^3$, while in the case of $\mathbb{S}^2$ we achieve this classification for all 6-vertex graphs. Additionally, by increasing the number of embeddings of selected graphs, we improve the previously known asymptotic lower bound on the maximum number of realizations.

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

5.12. Voronoï diagram of orthogonal polyhedra in two and three dimensions

Participants: Ioannis Emiris, Christina Katsamaki.

In [20], we study Voronoï diagrams, which are a fundamental geometric data structure for obtaining proximity relations. We consider collections of axis-aligned orthogonal polyhedra in two and three-dimensional space under the max-norm, which is a particularly useful scenario in certain application domains. We construct the exact Voronoï diagram inside an orthogonal polyhedron with holes defined by such polyhedra. Our approach avoids creating full-dimensional elements on the Voronoï diagram and yields a skeletal representation of the input object. We introduce a complete algorithm in 2D and 3D that follows the subdivision paradigm relying on a bounding-volume hierarchy; this is an original approach to the problem. The complexity is adaptive and comparable to that of previous methods. Under a mild assumption it is $O(n/D)$ in 2D or $O(na^2/D^2)$ in 3D, where n is the number of sites, namely edges or facets resp., D is the maximum cell size for the subdivision to stop, and a bounds vertex cardinality per facet. We also provide a numerically stable, open-source implementation in Julia, illustrating the practical nature of our algorithm.

The software was developed during Katsamaki's internship in 2018 at Sophia-Antipolis under the supervision of Bernard Mourrain. The problem has been proposed by our industrial collaborator ANSYS Hellas. The paper is based on Katsamaki's MSc thesis.

5.13. Near-Neighbor Preserving Dimension Reduction for Doubling Subsets of L_1

Participants: Ioannis Emiris, Ioannis Psarros.

In [21], we study randomized dimensionality reduction which has been recognized as one of the fundamental techniques in handling high-dimensional data. Starting with the celebrated Johnson-Lindenstrauss Lemma, such reductions have been studied in depth for the Euclidean (L_2) metric, but much less for the Manhattan (L_1) metric. Our primary motivation is the approximate nearest neighbor problem in L_1. We exploit its reduction to the decision-with-witness version, called approximate near neighbor, which incurs a roughly logarithmic overhead. In 2007, Indyk and Naor, in the context of approximate nearest neighbors, introduced the notion of nearest neighbor-preserving embeddings. These are randomized embeddings between two metric spaces with guaranteed bounded distortion only for the distances between a query point and a point set. Such embeddings are known to exist for both L_2 and L_1 metrics, as well as for doubling subsets of L_2. The case that remained open were doubling subsets of L_1. In this paper, we propose a dimension reduction by means of a near neighbor-preserving embedding for doubling subsets of L_1. Our approach is to represent the pointset with a carefully chosen covering set, then randomly project the latter. We study two types of covering sets: c-approximate r-nets and randomly shifted grids, and we discuss the tradeoff between them in terms of preprocessing time and target dimension. We employ Cauchy variables: certain concentration bounds derived should be of independent interest.

This is joint work with Vassilis Margonis (NKUA), and is based on his MSc thesis.

5.14. On the cross-sectional distribution of portfolio returns

Participants: Ioannis Emiris, Apostolos Chalkis.

The aim of the paper [24] is to study the distribution of portfolio returns across portfolios, and for given asset returns. We focus on the most common type of investment, considering portfolios whose weights are non-negative and sum up to 1. We provide algorithms and formulas from computational geometry and the literature on splines to compute the exact values of the probability density function, and of the cumulative distribution function, at any point. We also provide closed form solutions for the computation of its first four moments, and an algorithm to compute the higher moments. All algorithms and formulas allow also for equal asset returns.

This is a joint work with Ludovic Calès (JRC - European Commission - Joint Research Centre, Ispra).

5.15. Enumerating the morphologies of non-degenerate Darboux cyclides

Participant: Bernard Mourrain.

In [19] we provide an enumeration of all possible morphologies of non-degenerate Darboux cyclides. Based on the fact that every Darboux cyclide in $\mathbb{R}^3$ is the stereographic projection of the intersection surface of a sphere and a quadric in $\mathbb{R}^4$, we transform the enumeration problem of morphologies of Darboux cyclides to the enumeration of the algebraic sequences that characterize the intersection of a sphere and a quadric in $\mathbb{R}^4$.

This is a joint work with Mingyang Zhao, Xiaohong Jia (KLMM - Key Laboratory of Mathematics Mechanization, Beijing, China), Changhe Tu (Shandong University, China), Wenping Wang (Computer Graphics Group, Department of Computer Science, Hong Kong, China).

5.16. Anisotropic convolution surfaces

Participants: Alvaro Fuentes Suarez, Evelyne Hubert.

Convolution surfaces with 1D skeletons have been limited to close-to-circular normal sections. The new formalism and method presented in [10] allows for ellipsoidal normal sections. Anisotropy is prescribed on G^1 skeletal curves, chosen as circular splines, by a rotation angle and the three radii of an ellipsoid at each extremity. This lightweight model creates smooth shapes that previously required tweaking the skeleton or supplementing it with 2D pieces. The scale invariance of our formalism achieves excellent radii control and thus lends itself to approximate a variety of shapes. The construction of a scaffold is extended to skeletons with G^1 branches. It projects onto the convolution surface as a quad mesh with skeleton bound edge-flow.

This is a joint work with Cédric Zanni (MFX Inria NGE).

5.17. A non-iterative method for robustly computing the intersections between a line and a curve or surface

Participant: Laurent Busé.

The need to compute the intersections between a line and a high-order curve or surface arises in a large number of finite element applications. Such intersection problems are easy to formulate but hard to solve robustly. In [18], we introduce a non-iterative method for computing intersections by solving a matrix singular value decomposition (SVD) and an eigenvalue problem. That is, all intersection points and their parametric coordinates are determined in one-shot using only standard linear algebra techniques available in most software libraries. As a result, the introduced technique is far more robust than the widely used Newton-Raphson iteration or its variants. The maximum size of the considered matrices depends on the polynomial degree q of the shape functions and is $2q \times 3q$ for curves and $6q^2 \times 8q^2$ for surfaces. The method has its origin in algebraic geometry and has here been considerably simplified with a view to widely used high-order finite elements. In addition, the method is derived from a purely linear algebra perspective without resorting to algebraic geometry terminology. A complete implementation is available from http://bitbucket.org/nitro-project/.

This is joint work with Xiao Xiao and Fehmi Cirak (Cambridge, UK).

5.18. Cooperative Visual-Inertial Sensor Fusion: the Analytic Solution

Participant: Bernard Mourrain.

In [15], we analyze the visual–inertial sensor fusion problem in the cooperative case of two agents, and proves that this sensor fusion problem is equivalent to a simple polynomial equations system that consists of several linear equations and three polynomial equations of second degree. The analytic solution of this polynomial equations system is easily obtained by using an algebraic method. In other words, this letter provides the analytic solution to the visual–inertial sensor fusion problem in the case of two agents. The power of the analytic solution is twofold. From one side, it allows us to determine the relative state between the agents (i.e., relative position, speed, and orientation) without the need of an initialization. From another side, it provides fundamental insights into all the theoretical aspects of the problem. This letter mainly focuses on the first issue. However, the analytic solution is also exploited to obtain basic structural properties of the problem that characterize the observability of the absolute scale and the relative orientation. Extensive simulations and real experiments show that the solution is successful in terms of precision and robustness.

This is a joint work with Agostino Martinelli and Alexander Oliva (CHROMA, Inria Grenoble).

5.19. Overlapping Multi-Patch Structures in Isogeometric Analysis

Participant: Angelos Mantzaflaris.

In isogeometric analysis (IGA) the domain of interest is usually represented by B-spline or NURBS patches, as they are present in standard CAD models. Complex domains can often be represented as a union of simple overlapping subdomains, parameterized by (tensor-product) spline patches. Numerical simulation on such overlapping multi-patch domains is a serious challenge in IGA. To obtain non-overlapping subdomains one would usually reparameterize the domain or trim some of the patches. Alternatively, one may use methods that can handle overlapping subdomains. In [13] we propose a non-iterative, robust and efficient method defined directly on overlapping multi-patch domains. Consequently, the problem is divided into several sub-problems, which are coupled in an appropriate way. The resulting system can be solved directly in a single step. We compare the proposed method with iterative Schwarz domain decomposition approaches and observe that our method reduces the computational cost significantly, especially when handling subdomains with small overlaps. Summing up, our method significantly simplifies the domain parameterization problem, since we can represent any domain of interest as a union of overlapping patches without the need to introduce trimming curves/surfaces. The performance of the proposed method is demonstrated by several numerical experiments for the Poisson problem and linear elasticity in two and three dimensions.

This is a joint work with S. Kargaran, B. Jüttler, S. Kleiss and T. Takacs. (RICAM - Johann Radon Institute for Computational and Applied Mathematics and Institute of Applied Geometry, Linz, Austria)

5.20. First Order Error Correction for Trimmed Quadrature in Isogeometric Analysis

Participant: Angelos Mantzaflaris.

In [23] we develop a specialized quadrature rule for trimmed domains , where the trimming curve is given implicitly by a real-valued function on the whole domain. We follow an error correction approach: In a first step, we obtain an adaptive subdivision of the domain in such a way that each cell falls in a pre-defined base case. We then extend the classical approach of linear approximation of the trimming curve by adding an error correction term based on a Taylor expansion of the blending between the linearized implicit trimming curve and the original one. This approach leads to an accurate method which improves the convergence of the quadrature error by one order compared to piecewise linear approximation of the trimming curve. It is at the same time efficient, since essentially the computation of one extra one-dimensional integral on each trimmed cell is required. Finally, the method is easy to implement, since it only involves one additional line integral and refrains from any point inversion or optimization operations. The convergence is analyzed theoretically and numerical experiments confirm that the accuracy is improved without compromising the computational complexity.

This is joint work with B. Jüttler and F. Scholz. (Institute of Applied Geometry, Linz, Austria).

5.21. Consistent discretization of higher-order interface models for thin layers and elastic material surfaces, enabled by isogeometric cut-cell methods

Participant: Angelos Mantzaflaris.

Many interface formulations, e.g. based on asymptotic thin interphase models or material surface theories, involve higher-order differential operators and discontinuous solution fields. In [11] we are taking first steps towards a variationally consistent discretization framework that naturally accommodates these two challenges by synergistically combining recent developments in isogeometric analysis and cut-cell finite element methods. Its basis is the mixed variational formulation of the elastic interface problem that provides access to jumps in displacements and stresses for incorporating general interface conditions. Upon discretization with smooth splines, derivatives of arbitrary order can be consistently evaluated, while cut-cell meshes enable discontinuous solutions at potentially complex interfaces. We demonstrate via numerical tests for three specific nontrivial interfaces (two regimes of the Benveniste–Miloh classification of thin layers and the Gurtin–Murdoch material surface model) that our framework is geometrically flexible and provides optimal higher-order accuracy in the bulk and at the interface.

This is joint work with Zhilin Han, Changzheng Cheng, (HFUT - Hefei University of Technology, China), Chien-Ting Wu, S. Stoter, S. Mogilevskaya, and D. Schillinger (Department of Civil, Environmental and Geo-Engineering, University of Minnesota, USA).

5.22. Design of Self-Supporting Surfaces with Isogeometric Analysis

Participant: Angelos Mantzaflaris.

Self-supporting surfaces are widely used in contemporary architecture, but their design remains a challenging problem. This paper aims to provide a heuristic strategy for the design of complex self-supporting surfaces. In our method, presented in [17] non-uniform rational B-spline (NURBS) surfaces are used to describe the smooth geometry of the self-supporting surface. The equilibrium state of the surface is derived with membrane shell theory and Airy stresses within the surfaces are used as tunable variables for the proposed heuristic design strategy. The corresponding self-supporting shapes to the given stress states are calculated by the nonlinear isogeometric analysis (IGA) method. Our validation using analytic catenary surfaces shows that the proposed method finds the correct self-supporting shape with a convergence rate one order higher than the degree of the applied NURBS basis function. Tests on boundary conditions show that the boundary's influence propagates along the main stress directions in the surface. Various self-supporting masonry structures, including models with complex topology, are constructed using the presented method. Compared with existing methods such as thrust network analysis and dynamic relaxation, the proposed method benefits from the advantages of NURBS-based IGA, featuring smooth geometric description, good adaption to complex shapes and increased efficiency of computation.

This is joint work with Yang Xia, Ping Hu (Dalian University of Technology, China), Bert Jüttler (Institute of Applied Geometry, Linz, Austria), Hao Pan (Microsoft Research Asia, China), Wenping Wang (CSE - Department of Computer Science and Engineering, HKUST, Honk Kong, China).

5.23. Low-rank space-time decoupled isogeometric analysis for parabolic problems with varying coefficients

Participant: Angelos Mantzaflaris.

In [14] we present a space-time isogeometric analysis scheme for the discretization of parabolic evolution equations with diffusion coefficients depending on both time and space variables. The problem is considered in a space-time cylinder in $\mathbb{R}^{d+1}$, with $d = 2, 3$ and is discretized using higher-order and highly-smooth spline spaces. This makes the matrix formation task very challenging from a computational point of view. We overcome this problem by introducing a low-rank decoupling of the operator into space and time components. Numerical experiments demonstrate the efficiency of this approach.

This work was done jointly with F. Scholz and I. Toulopoulos (RICAM - Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria).

CARAMBA Project-Team

7. New Results

7.1. Algebraic Curves for Cryptology

7.1.1. Cocks-Pinch Curves of Embedding Degrees Five to Eight and Optimal Ate Pairing Computation

Participants: Aurore Guillevic, Simon Masson, Emmanuel Thomé.

In [21] we explored a modification of the Cocks-Pinch method to generate pairing-friendly curves resistant to the Special-Tower-NFS algorithm (STNFS). We carefully estimated the cost of the STNFS attack for existing families of curves, and chose curves of embedding degree five to eight. For prime embedding degrees 5 and 7, our curves are naturally immune to the STNFS attack, but their performance level is not high. For composite embedding degrees 6 and 8 for which the TNFS attack applies, we chose the parameters from a family that is general enough to thwart the “special” variant STNFS; we also optimized these parameter choices so that these curves can have a reasonably efficient pairing computation, close with the very best possible curve choices.

7.1.2. A Short-List of Pairing-Friendly Curves Resistant to Special TNFS at the 128-bit Security Level

Participant: Aurore Guillevic.

The preprint [20] applies the refinements of the paper [22] to estimate the cost of the Special Tower NFS algorithm for particular pairing-friendly curves, whose target group is $\mathbb{F}_{p^n}$, and where the characteristic is special, parameterized by a low degree polynomial. We show that with a new variant of the polynomial selection, the estimated cost is reduced, but stays above the theoretical bound of the Special NFS $L_{p^n}(1/3,(32/9)^{1/3})$. This variant does not apply to the Cocks-Pinch curves of [21]. We list nine interesting pairing-friendly curves of embedding degrees between 10 and 16 at the 128-bit security level.

7.1.3. A Practical Attack on ECDSA Implementations Using wNAF Representation

Participants: Gabrielle de Micheli, Cécile Pierrot, Rémi Piau.

ECDSA is a widely deployed public key signature protocol that uses elliptic curves. One way of attacking ECDSA with wNAF implementation for the scalar multiplication is to perform a side-channel analysis to collect information, then use a lattice based method to recover the secret key. In [18], we re-investigate the construction of the lattice used in one of these methods, the Extended Hidden Number Problem (EHNP). We find the secret key with only 3 signatures, thus reaching the theoretical bound never achieved before. Our attack is more efficient than previous attacks, has better probability of success, and is still able to find the secret key with a small amount of erroneous traces, up to 2% of false digits.

7.1.4. Algorithmic Aspects of Elliptic Bases in Finite Field Discrete Logarithm Algorithms

Participant: Cécile Pierrot.

Elliptic bases give an elegant way of representing finite field extensions and were used as a starting point for small characteristic finite field discrete logarithm algorithms. This idea has been proposed by two groups, in order to achieve provable quasi-polynomial time algorithms for computing discrete logarithms in small characteristic finite fields. In [23], together with Antoine Joux, we do not try to achieve a provable algorithm, but instead we investigate the practicality of heuristic algorithms based on elliptic bases.

7.1.5. A Fast Randomized Geometric Algorithm for Computing Riemann-Roch Spaces

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

In [7], we proposed a probabilistic 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. Most of the results of this work have been obtained in 2018. In 2019, we have strengthened these results and revised the associated paper. This new version of the paper has been accepted for publication in the journal Mathematics of Computation.

7.1.6. Counting Points on Hyperelliptic Curves

Participants: Pierrick Gaudry, Pierre-Jean Spaenlehauer.

Two works with Simon Abelard [1], [2] following his PhD thesis about improved complexities for counting point algorithms of hyperelliptic curves with or without real multiplication are now formally published as journal articles.

7.1.7. Verifiable Delay Functions from Supersingular Isogenies and Pairings

Participant: Simon Masson.

Together with Luca De Feo, Christophe Petit and Antonio Sanso, we introduce in [11] two verifiable delay functions based on isogenies of supersingular elliptic curves and pairing. We discuss both the advantages and drawbacks of our constructions, we study their security and we demonstrate their practicality with a proof-of-concept implementation. This work appears in the proceedings of ASIACRYPT'2019.

7.1.8. Isogeny Graphs With Maximal Real Multiplication

Participant: Emmanuel Thomé.

Emmanuel Thomé and Sorina Ionica (post-doctoral fellow in the former CARAMEL team in 2012) worked on a new algorithm for computing isogeny graphs for Jacobians of curves having the special property that the intersection of their endomorphism ring with its real subfield is maximal. The resulting algorithm is the first depth-first algorithm for this task. The work [6] was finally published.

7.2. The Number Field Sieve – High-Level Results

7.2.1. A New Ranking Function for Polynomial Selection in the Number Field Sieve

Participant: Paul Zimmermann.

With Nicolas David (ÉNS Paris-Saclay, France), we designed a new ranking function for polynomial selection in the Number Field Sieve. The previous ranking function was only considering the *mean* of the so-called α-value, which measures how small primes divide the norm of the polynomial. The new function also takes into account the *variance* of the corresponding distribution. This partially explains why the previous function did sometimes fail to correctly identify the best polynomials. The new ranking function is implemented in Cado-NFS (branch `dist-alpha`) and is detailed in [3].

7.2.2. On the Alpha Value of Polynomials in the Tower Number Field Sieve Algorithm

Participant: Aurore Guillevic.

With Shashank Singh from IISER Bhopal (former post-doc at CARAMBA in 2017), we generalized the ranking function α for the Tower setting of the Number Field Sieve in [22]. In the relation collection of the NFS algorithm, one tests the smoothness of algebraic norms (computed with resultants). The α function measures the bias of the average valuation at small primes of algebraic norms, compared to the average valuation at random integers of the same size. A negative α means more small divisors than average. We then estimate the total number of relations with a Monte-Carlo simulation, as a generalized Murphy's E function, and finally give a rough estimate of the total cost of TNFS for finite fields $\mathbb{F}_{p^k}$ of popular pairing-friendly curves.

7.2.3. Faster Individual Discrete Logarithms in Finite Fields of Composite Extension Degree

Participant: Aurore Guillevic.

We improved the previous work [30] 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 was finally published in 2019 [4].

7.3. The Number Field Sieve – Implementation Results

7.3.1. Parallel Structured Gaussian Elimination for the Number Field Sieve

Participant: Paul Zimmermann.

Together with Charles Bouillaguet (University of Lille, France), we completely re-designed the structured Gaussian elimination step of Cado-NFS (called `merge`). The new algorithm is fully parallel, and scales quite well. With 32 cores on modern hardware, the `merge`-step of RSA-512 (factored in 1999) now takes only 20 seconds, and for the hidden SNFS DLP-1024 record (done in 2017) it takes only 140 seconds [16].

7.4. Computer Arithmetic

7.4.1. Breaking Randomized Mixed-Radix Scalar Multiplication Algorithms

Participant: Jérémie Detrey.

Together with Laurent Imbert (LIRMM, France), we designed in [13] an attack against a recently published randomized elliptic-curve scalar multiplication scheme based on covering systems of congruences. We also proposed a more robust algorithm based on a mixed-radix representation of the scalar. However, under strong security hypotheses, this algorithm may still allow a virtual powerful attacker to recover much more information than what was first expected. This led us to the conclusion that randomized algorithms based on the mixed-radix number system should be avoided.

7.5. Symmetric Cryptology

7.5.1. Vectorial Boolean Functions with Very Low Differential-Linear Uniformity Using Maiorana-McFarland Type Construction

Participant: Bimal Mandal.

With Deng Tang and Subhamoy Maitra, we constructed in [14] a new class of balanced vectorial Boolean functions with very low differential-linear uniformity, whose coordinate functions are derived by modifying the Maiorana–McFarland bent functions. Further, we provided a combinatorial count of hardware gates required to implement such circuits.

7.5.2. Analysis of Boolean Functions in a Restricted (Biased) Domain

Participant: Bimal Mandal.

This work with Subhamoy Maitra, Thor Martinsen, Dibyendu Roy and Pantelimon Stanica [8] is a substantially revised and extended version of the paper “Tools in analyzing linear approximation for Boolean functions related to FLIP” that appeared in the proceedings of Indocrypt 2018 [32]. We proposed a technique to study the cryptographic properties of Boolean functions, whose inputs do not follow uniform distribution, and obtain a lower bound for the bias of the nonlinear filter function of FLIP by using biased Walsh–Hadamard transform. Our results provided more accurate calculation of the biases of Boolean function over restricted domain, which help to determine the security parameter of FLIP type ciphers.

7.5.3. Forkcipher: a New Primitive for Authenticated Encryption of Very Short Messages

Participant: Virginie Lallemand.

Together with Elena Andreeva, Antoon Purnal, Reza Reyhanitabar, Arnab Roy and Damian Vizár, we proposed a candidate to the NIST Lightweight competition that we also published at Asiacrypt 2019 [10]. Our proposal is based on the so-called forkcipher construction that was previously presented and investigated by a subset of the authors and which provides authenticated encryption optimized for short messages. Our NIST candidate is called ForkAE, and as required by NIST it is based on well investigated primitives, out of which the Skinny tweakable cipher. ForkAE is one of the 32 candidates that were selected to continue to Round 2 out of 56 valid submissions.

7.5.4. Computing AES Related-Key Differential Characteristics With Constraint Programming

Participant: Marine Minier.

In [5], with David Gérault, Pascal Lafourcade, and Christine Solnon, we improve existing Constraint Programming (CP) approaches for computing optimal related-key differential characteristics: we add new constraints that detect inconsistencies sooner, and we introduce a new decomposition of the problem in two steps. These improvements allow us to compute all optimal related-key differential characteristics for AES-128, AES-192 and AES-256 in a few hours.

7.5.5. Participation in the NIST Lightweight Cryptography Standardization Process

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

The team is actively taking part in the lightweight cryptography standardization process of the NIST. The two major actions that have been taken are the following:

- Proposition of two candidates, namely Lilliput-AE (Alexandre Adomnicai, Thierry P. Berger, Christophe Clavier, Julien Francq, Paul Huynh, Virginie Lallemand, Kévin LeGouguec, Marine Minier, Léo Reynaud and Gaël Thomas) and ForkAE (Elena Andreeva, Virginie Lallemand, Antoon Purnal, Reza Reyhanitabar, Arnab Roy and Damian Vizár). ForkAE made it to the second round, but unfortunately a weak point has been detected in the design of Lilliput-AE.
- Organization of regular cryptanalysis meetings with other french cryptographers. Since the publication of the 56 proposals, four meetings have been held and some tangible results have already been achieved. As an example, the meeting participants found a practical differential forgery attack against the proposal named *Quartet*. The details have been made public on the NIST mailing list and they made the NIST remove this candidate from consideration.

7.5.6. Cryptanalysis of SKINNY in the Framework of the SKINNY 2018-2019 Cryptanalysis Competition

Participant: Virginie Lallemand.

Together with Patrick Derbez (University of Rennes) and Aleksei Udovenko (University of Luxembourg) we investigated in [12] the security of the SKINNY tweakable block cipher, a lightweight symmetric cipher proposed at Crypto in 2016. Our setting was the one of the SKINNY 2018-2019 Cryptanalysis Competition, that is we looked for attacks that can be run in practical time and that succeed with a data set reduced to the provided set of 2^{20} (plaintext, ciphertext). We solved the challenges (meaning that we experimentally recovered the 128-bit key) for up to 10-round SKINNY-128-128 and 12-round SKINNY-64-128. To this day these are the best results reported in this setting.

7.6. E-voting

7.6.1. Belenios: a Simple Private and Verifiable Electronic Voting System

Participant: Pierrick Gaudry.

In [9], written with Véronique Cortier and Stéphane Glondu, we have summarized the current state of our voting platform Belenios. It was the occasion to put in a single place the description of several sub-parts of the protocol that are otherwise spread in many articles. We also made statistics regarding the use of the platform during the year 2018, and discussed how security features were or were not activated by the users.

7.6.2. A Simple Alternative to Benaloh Challenge for the Cast-as-Intended Property in Helios/Belenios

Participant: Pierrick Gaudry.

In a short note [17] written with Véronique Cortier, Jannik Dreier, and Mathieu Turuani from the PESTO team, we propose a simple technique that can be added to an Helios-like e-voting protocol, so that the voter can check whether their potentially infected computer has not silently changed their vote.

7.6.3. Breaking the Encryption Scheme of the Moscow Internet Voting System

Participant: Pierrick Gaudry.

In [19], written in collaboration with Alexander Golovnev (Harvard), we explain the vulnerabilities we have found in an Internet voting system used for the election for the representatives of the Moscow Duma that took place in September 2019. The weaknesses in the encryption scheme (based on the discrete logarithm problem in finite fields) were found in the source code that was made available in July 2019 as part of a public testing.

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
- Attribute and predicate encryption schemes
- New primitives for efficient anonymous authentication
- Applications to machine learning
- Blockchain protocols
- Searchable Encryption

DATASHAPE Project-Team

5. New Results

5.1. Algorithmic aspects of topological and geometric data analysis

5.1.1. Sampling and Meshing Submanifolds

Participants: Jean-Daniel Boissonnat, Siargey Kachanovich.

In collaboration with Mathijs Wintraecken (IST Autria).

This work [41], [11] presents a rather simple tracing algorithm to sample and mesh an m-dimensional submanifold of $\mathbb{R}^d$ for arbitrary m and d. We extend the work of Dobkin et al. to submanifolds of arbitrary dimension and codimension. The algorithm is practical and has been thoroughly investigated from both theoretical and experimental perspectives. The paper provides a full description and analysis of the data structure and of the tracing algorithm. The main contributions are : 1. We unify and complement the knowledge about Coxeter and Freudenthal-Kuhn triangulations. 2. We introduce an elegant and compact data structure to store Coxeter or Freudenthal-Kuhn triangulations and describe output sensitive algorithms to compute faces and cofaces or any simplex in the triangulation. 3. We present a manifold tracing algorithm based on the above data structure. We provide a detailled complexity analysis along with experimental results that show that the algorithm can handle cases that are far ahead of the state-of-the-art.

5.1.2. Topological correctness of PL-approximations of isomanifolds

Participant: Jean-Daniel Boissonnat.

In collaboration with Mathijs Wintraecken (IST Autria).

Isomanifolds are the generalization of isosurfaces to arbitrary dimension and codimension, i.e. manifolds defined as the zero set of some multivariate multivalued function $f : \mathbb{R}^d \rightarrow \mathbb{R}^{d-n}$. A natural (and efficient) way to approximate an isomanifold is to consider its Piecewise-Linear (PL) approximation based on a triangulation $\mathcal{T}$ of the ambient space $\mathbb{R}^d$. In this paper [43], we give conditions under which the PL-approximation of an isomanifold is topologically equivalent to the isomanifold. The conditions are easy to satisfy in the sense that they can always be met by taking a sufficiently fine triangulation $\mathcal{T}$. This contrasts with previous results on the triangulation of manifolds where, in arbitrary dimensions, delicate perturbations are needed to guarantee topological correctness, which leads to strong limitations in practice. We further give a bound on the Fréchet distance between the original isomanifold and its PL-approximation. Finally we show analogous results for the PL-approximation of an isomanifold with boundary.

5.1.3. Dimensionality Reduction for k-Distance Applied to Persistent Homology

Participants: Jean-Daniel Boissonnat, Kunal Dutta.

In collaboration with Shreya Arya (Duke University)

Given a set P of n points and a constant k, we are interested in computing the persistent homology of the Čech filtration of P for the k-distance, and investigate the effectiveness of dimensionality reduction for this problem, answering an open question of Sheehy [*Proc. SoCG, 2014*] [38]. We first show using the Johnson-Lindenstrauss lemma, that the persistent homology can be preserved up to a $(1 \pm \epsilon)$ factor while reducing dimensionality to $O(k \log n/\varepsilon^2)$. Our main result shows that the target dimension can be improved to $O(\log n/\varepsilon^2)$ under a reasonable and naturally occuring condition. The proof involves a multi-dimensional variant of the Hanson-Wright inequality for subgaussian quadratic forms and works when the random matrices are used for the Johnson-Lindenstrauss mapping are subgaussian. This includes the Gaussian matrices of Indyk-Motwani, the sparse random matrices of Achlioptas and the Ailon-Chazelle fast Johnson-Lindenstrauss transform. To provide evidence that our condition encompasses quite general situations, we show that it is satisfied when the points are independently distributed (i) in $\mathbb{R}^D$ under a subgaussian distribution, or (ii) on a spherical shell in $\mathbb{R}^D$ with a minimum angular separation, using Gershgorin's theorem. Our results also show that the JL-mapping preserves up to a $(1 \pm \epsilon)$ factor, the Rips and Delaunay filtrations for the k-distance, as well as the Čech filtration for the approximate k-distance of Buchet et al.

5.1.4. Edge Collapse and Persistence of Flag Complexes

Participants: Jean-Daniel Boissonnat, Siddharth Pritam.

In this article [42], we extend the notions of dominated vertex and strong collapse of a simplicial complex as introduced by J. Barmak and E. Miniam adn build on the initial success of [30]. We say that a simplex (of any dimension) is dominated if its link is a simplicial cone. Domination of edges appear to be very powerful and we study it in the case of flag complexes in more detail. We show that edge collapse (removal of dominated edges) in a flag complex can be performed using only the 1-skeleton of the complex. Furthermore, the residual complex is a flag complex as well. Next we show that, similar to the case of strong collapses, we can use edge collapses to reduce a flag filtration $\mathcal{F}$ to a smaller flag filtration $\mathcal{F}^c$ with the same persistence. Here again, we only use the 1-skeletons of the complexes. The resulting method to compute $\mathcal{F}^c$ is simple and extremely efficient and, when used as a preprocessing for Persistence Computation, leads to gains of several orders of magnitude wrt the state-of-the-art methods (including our previous approach using strong collapse). The method is exact, irrespective of dimension, and improves performance of Persistence Computation even in low dimensions. This is demonstrated by numerous experiments on publicly available data.

5.1.5. DTM-based Filtrations

Participants: Frédéric Chazal, Marc Glisse, Raphael Tinarrage.

In collaboration with Anai, Hirokazu and Ike, Yuichi and Inakoshi, Hiroya and Umeda, Yuhei (Fujitsu Labs).

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 [15], 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.

5.1.6. Recovering the homology of immersed manifolds

Participant: Raphael Tinarrage.

Given a sample of an abstract manifold immersed in some Euclidean space, in [57], we describe a way to recover the singular homology of the original manifold. It consists in estimating its tangent bundle -seen as subset of another Euclidean space- in a measure theoretic point of view, and in applying measure-based filtrations for persistent homology. The construction we propose is consistent and stable, and does not involve the knowledge of the dimension of the manifold.

5.1.7. Regular triangulations as lexicographic optimal chains

Participant: David Cohen-Steiner.

In collaboration with André Lieutier and Julien Vuillamy (Dassault Systèmes).

We introduce [46] a total order on n-simplices in the n-Euclidean space for which the support of the lexicographic-minimal chain with the convex hull boundary as boundary constraint is precisely the n-dimensional Delaunay triangulation, or in a more general setting, the regular triangulation of a set of weighted points. This new characterization of regular and Delaunay triangulations is motivated by its possible generalization to submanifold triangulations as well as the recent development of polynomial-time triangulation algorithms taking advantage of this order.

5.1.8. 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 [35].

5.1.9. Computing Persistent Homology with Various Coefficient Fields in a Single Pass

Participants: Jean-Daniel Boissonnat, Clément Maria.

This article [18] introduces an algorithm to compute the persistent homology of a filtered complex with various coefficient fields in a single matrix reduction. The algorithm is output-sensitive in the total number of distinct persistent homological features in the diagrams for the different coefficient fields. This computation allows us to infer the prime divisors of the torsion coefficients of the integral homology groups of the topological space at any scale, hence furnishing a more informative description of topology than persistence in a single coefficient field. We provide theoretical complexity analysis as well as detailed experimental results. The code is part of the Gudhi software library.

5.1.10. 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 [31] 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.

5.1.11. Decomposition of exact pfd persistence bimodules

Participant: Steve Oudot.

In collaboration with Jérémy Cochoy (Symphonia).

In [24] we identify a certain class of persistence modules indexed over $\mathbb{R}^2$ that are decomposable into direct sums of indecomposable summands called blocks. The conditions on the modules are that they are both pointwise finite-dimensional (pfd) and exact. Our proof follows the same scheme as the one for pfd persistence modules indexed over $\mathbb{R}$, yet it departs from it at key stages due to the product order not being a total order on $\mathbb{R}^2$, which leaves some important gaps open. These gaps are filled in using more direct arguments. Our work is motivated primarily by the study of interlevel-sets persistence, although the proposed results reach beyond that setting.

5.1.12. Level-sets persistence and sheaf theory

Participants: Nicolas Berkouk, Steve Oudot.

In collaboration with Grégory Ginot (Paris 13).

In [39] we provide an explicit connection between level-sets persistence and derived sheaf theory over the real line. In particular we construct a functor from 2-parameter persistence modules to sheaves over R, as well as a functor in the other direction. We also observe that the 2-parameter persistence modules arising from the level sets of Morse functions carry extra structure that we call a Mayer-Vietoris system. We prove classification, barcode decomposition, and stability theorems for these Mayer-Vietoris systems, and we show that the aforementioned functors establish a pseudo-isometric equivalence of categories between derived constructible sheaves with the convolution or (derived) bottleneck distance and the interleaving distance of strictly pointwise finite-dimensional Mayer-Vietoris systems. Ultimately, our results provide a functorial equivalence between level-sets persistence and derived pushforward for continuous real-valued functions.

5.1.13. Intrinsic Interleaving Distance for Merge Trees

Participant: Steve Oudot.

In collaboration with Ellen Gasparovic (Union College), Elizabeth Munch (Michigan State), Katharine Turner (Australian National University), Bei Wang (Utah), and Yusu Wang (Ohio-State).

Merge trees are a type of graph-based topological summary that tracks the evolution of connected components in the sublevel sets of scalar functions. They enjoy widespread applications in data analysis and scientific visualization. In [49] we consider the problem of comparing two merge trees via the notion of interleaving distance in the metric space setting. We investigate various theoretical properties of such a metric. In particular, we show that the interleaving distance is intrinsic on the space of labeled merge trees and provide an algorithm to construct metric 1-centers for collections of labeled merge trees. We further prove that the intrinsic property of the interleaving distance also holds for the space of unlabeled merge trees. Our results are a first step toward performing statistics on graph-based topological summaries.

5.2. Statistical aspects of topological and geometric data analysis

5.2.1. Estimating the Reach of a Manifold

Participants: Frédéric Chazal, Jisu Kim, Bertrand Michel.

In collaboration with E. Aamari (Univ. Paris-Diderot), A. Rinaldo, L. Wasserman (Carnegie Mellon University).

In [13], 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. This paper 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 an oracle framework where tangent spaces are known, and bounds assessing its efficiency are derived. In the case of i.i.d. random point cloud X_n, $\widehat{\tau}(X_n)$ is showed to achieve uniform expected loss bounds over a C^3-like model. Finally, we obtain upper and lower bounds on the minimax rate for estimating the reach.

5.2.2. A statistical test of isomorphism between metric-measure spaces using the distance-to-a-measure signature

Participant: Claire Brecheteau.

In [20], we introduce the notion of DTM-signature, a measure on $\mathbb{R}$ that can be associated to any metric-measure space. This signature is based on the function distance to a measure (DTM) introduced in 2009 by Chazal, Cohen-Steiner and Mérigot. It leads to a pseudo-metric between metric-measure spaces, that is bounded above by the Gromov-Wasserstein distance. This pseudo-metric is used to build a statistical test of isomorphism between two metric-measure spaces, from the observation of two N-samples.

The test is based on subsampling methods and comes with theoretical guarantees. It is proven to be of the correct level asymptotically. Also, when the measures are supported on compact subsets of $\mathbb{R}^d$, rates of convergence are derived for the $L1$-Wasserstein distance between the distribution of the test statistic and its subsampling approximation. These rates depend on some parameter $\rho > 1$. In addition, we prove that the power is bounded above by $\exp(-CN1/\rho)$, with C proportional to the square of the aforementioned pseudo-metric between the metric-measure spaces. Under some geometrical assumptions, we also derive lower bounds for this pseudo-metric.

An algorithm is proposed for the implementation of this statistical test, and its performance is compared to the performance of other methods through numerical experiments.

5.2.3. 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 [21], 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 $\mathbb{X}_n$ is an i.i.d. sample from a density on $[0,1]^d$. For the Č ech 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 persistences of such diagrams. Those two approaches (stability and convergence) 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 $\alpha \geq d$.

5.2.4. Understanding the Topology and the Geometry of the Persistence Diagram Space via Optimal Partial Transport

Participants: Vincent Divol, Théo Lacombe.

Despite the obvious similarities between the metrics used in topological data analysis and those of optimal transport, an optimal-transport based formalism to study persistence diagrams and similar topological descriptors has yet to come. In [48], by considering the space of persistence diagrams as a measure space, and by observing that its metrics can be expressed as solutions of optimal partial transport problems, we introduce a generalization of persistence diagrams, namely Radon measures supported on the upper half plane. Such measures naturally appear in topological data analysis when considering continuous representations of persistence diagrams (e.g. persistence surfaces) but also as limits for laws of large numbers on persistence diagrams or as expectations of probability distributions on the persistence diagrams space. We study the topological properties of this new space, which will also hold for the closed subspace of persistence diagrams. New results include a characterization of convergence with respect to transport metrics, the existence of Fréchet means for any distribution of diagrams, and an exhaustive description of continuous linear representations of persistence diagrams. We also showcase the usefulness of this framework to study random persistence diagrams by providing several statistical results made meaningful thanks to this new formalism.

5.3. Topological approach for multimodal data processing

5.3.1. A General Neural Network Architecture for Persistence Diagrams and Graph Classification

Participants: Frédéric Chazal, Théo Lacombe, Martin Royer.

In collaboration with Mathieu Carrière (Colombia Univ.) and Umeda Yuhei and Ike Yiuchi (Fujitsu Labs).

Persistence diagrams, the most common descriptors of Topological Data Analysis, encode topological properties of data and have already proved pivotal in many different applications of data science. However, since the (metric) space of persistence diagrams is not Hilbert, they end up being difficult inputs for most Machine Learning techniques. To address this concern, several vectorization methods have been put forward that embed persistence diagrams into either finite-dimensional Euclidean space or (implicit) infinite dimensional Hilbert space with kernels. In [44], we focus on persistence diagrams built on top of graphs. Relying on extended persistence theory and the so-called heat kernel signature, we show how graphs can be encoded by (extended) persistence diagrams in a provably stable way. We then propose a general and versatile framework for learning vectorizations of persistence diagrams, which encompasses most of the vectorization techniques used in the literature. We finally showcase the experimental strength of our setup by achieving competitive scores on classification tasks on real-life graph datasets.

5.3.2. Topological Data Analysis for Arrhythmia Detection through Modular Neural Networks

Participant: Frédéric Chazal.

In collaboration with Umeda Yuhei and Meryll Dindin (Fujitsu Labs).

In [47], we present an innovative and generic deep learning approach to monitor heart conditions from ECG signals.We focus our attention on both the detection and classification of abnormal heartbeats, known as arrhythmia. We strongly insist on generalization throughout the construction of a deep-learning model that turns out to be effective for new unseen patient. The novelty of our approach relies on the use of topological data analysis as basis of our multichannel architecture, to diminish the bias due to individual differences. We show that our structure reaches the performances of the state-of-the-art methods regarding arrhythmia detection and classification.

5.3.3. ATOL: Automatic Topologically-Oriented Learning

Participants: Frédéric Chazal, Martin Royer.

In collaboration with Umeda Yuhei and Ike Yiuchi (Fujitsu Labs).

There are abundant cases for using Topological Data Analysis (TDA) in a learning context, but robust topological information commonly comes in the form of a set of persistence diagrams, objects that by nature are uneasy to affix to a generic machine learning framework. In [56], we introduce a vectorisation method for diagrams that allows to collect information from topological descriptors into a format fit for machine learning tools. Based on a few observations, the method is learned and tailored to discriminate the various important plane regions a diagram is set into. With this tool one can automatically augment any sort of machine learning problem with access to a TDA method, enhance performances, construct features reflecting underlying changes in topological behaviour. The proposed methodology comes with only high level tuning parameters such as the encoding budget for topological features. We provide an open-access, ready-to-use implementation and notebook. We showcase the strengths and versatility of our approach on a number of applications. From emulous and modern graph collections to a highly topological synthetic dynamical orbits data, we prove that the method matches or beats the state-of-the-art in encoding persistence diagrams to solve hard problems. We then apply our method in the context of an industrial, difficult time-series regression problem and show the approach to be relevant.

5.3.4. Inverse Problems in Topological Persistence: a Survey

Participant: Steve Oudot.

In collaboration with Elchanan Solomon (Duke).

In [27] 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 rightinverses, and the second half focuses on injectivity, i.e. left inverses. Throughout, we highlightthe tools and theorems that underlie these advances, and direct the reader's attention to openproblems, both theoretical and applied.

5.3.5. *Intrinsic Topological Transforms via the Distance Kernel Embedding*

Participants: Clément Maria, Steve Oudot.

In collaboration with Elchanan Solomon (Duke).

Topological transforms are parametrized families of topological invariants, which, by analogy with transforms in signal processing, are much more discriminative than single measurements. The first two topological transforms to be defined were the Persistent Homology Transform and Euler Characteristic Transform, both of which apply to shapes embedded in Euclidean space. The contribution of this work [54] is to define topological transforms that depend only on the intrinsic geometry of a shape, and hence are invariant to the choice of embedding. To that end, given an abstract metric measure space, we define an integral operator whose eigenfunctions are used to compute sublevel set persistent homology. We demonstrate that this operator, which we call the distance kernel operator, enjoys desirable stability properties, and that its spectrum and eigenfunctions concisely encode the large-scale geometry of our metric measure space. We then define a number of topological transforms using the eigenfunctions of this operator, and observe that these transforms inherit many of the stability and injectivity properties of the distance kernel operator.

5.3.6. *A Framework for Differential Calculus on Persistence Barcodes*

Participant: Steve Oudot.

In collaboration with Jacob Leygonie and Ulrike Tillmann (Oxford).

In [52], we define notions of differentiability for maps from and to the space of persistence barcodes. Inspired by the theory of diffeological spaces, the proposed framework uses lifts to the space of ordered barcodes, from which derivatives can be computed. The two derived notions of differentiability (respectively from and to the space of barcodes) combine together naturally to produce a chain rule that enables the use of gradient descent for objective functions factoring through the space of barcodes. We illustrate the versatility of this framework by showing how it can be used to analyze the smoothness of various parametrized families of filtrations arising in topological data analysis.

5.4. Experimental research and software development

5.4.1. *Robust Stride Detector from Ankle-Mounted Inertial Sensors for Pedestrian Navigation and Activity Recognition with Machine Learning Approaches*

Participants: Bertrand Beaufils, Frédéric Chazal, Bertrand Michel.

In collaboration with Marc Grelet (Sysnav).

In [16], a stride detector algorithm combined with a technique inspired by zero velocity update (ZUPT) is proposed to reconstruct the trajectory of a pedestrian from an ankle-mounted inertial device. This innovative approach is based on sensor alignment and machine learning. It is able to detect 100% of both normal walking strides and more than 97% of atypical strides such as small steps, side steps, and backward walking that existing methods can hardly detect. This approach is also more robust in critical situations, when for example the wearer is sitting and moving the ankle or when the wearer is bicycling (less than two false detected strides per hour on average). As a consequence, the algorithm proposed for trajectory reconstruction achieves much better performances than existing methods for daily life contexts, in particular in narrow areas such as in a house. The computed stride trajectory contains essential information for recognizing the activity (atypical stride, walking, running, and stairs). For this task, we adopt a machine learning approach based on descriptors of these trajectories, which is shown to be robust to a large of variety of gaits. We tested our algorithm on recordings of healthy adults and children, achieving more than 99% success. The algorithm also achieved more than 97by children suffering from movement disorders. Compared to most algorithms in the literature, this original method does not use a fixed-size sliding window but infers this last in an adaptive way

5.4.2. *Robust pedestrian trajectory reconstruction from inertial sensor*

Participants: Bertrand Beaufils, Frédéric Chazal, Bertrand Michel.

In collaboration with Marc Grelet (Sysnav).

In [28], a strides detection algorithm combined with a technique inspired by Zero Velocity Update (ZUPT) is proposed using inertial sensors worn on the ankle. This innovative approach based on a sensors alignment and machine learning can detect both normal walking strides and atypical strides such as small steps, side steps and backward walking that existing methods struggle to detect. As a consequence, the trajectory reconstruction achieves better performances in daily life contexts for example, where a lot of these kinds of strides are performed in narrow areas such as in a house. It is also robust in critical situations, when for example the wearer is sitting and moving the ankle or bicycling, while most algorithms in the literature would wrongly detect strides and produce error in the trajectory reconstruction by generating movements.Our algorithm is evaluated on more than 7800 strides from seven different subjects performing several activities. We validated the trajectory reconstruction during motion capture sessions by analyzing the stride length. Finally, we tested the algorithm in a challenging situation by plotting the computed trajectory on the building map of an 5 hours and 30 minutes office worker recording.

5.5. Algorithmic and Combinatorial Aspects of Low Dimensional Topology

5.5.1. Treewidth, crushing and hyperbolic volume

Participant: Clément Maria.

In collaboration with Jessica S. Purcell (Monash University, Australia)

The treewidth of a 3-manifold triangulation plays an important role in algorithmic 3-manifold theory, and so it is useful to find bounds on the tree-width in terms of other properties of the manifold. In [26], we prove that there exists a universal constant c such that any closed hyperbolic 3-manifold admits a triangulation of tree-width at most the product of c and the volume. The converse is not true: we show there exists a sequence of hyperbolic 3-manifolds of bounded tree-width but volume approaching infinity. Along the way, we prove that crushing a normal surface in a triangulation does not increase the carving-width, and hence crushing any number of normal surfaces in a triangulation affects tree-width by at most a constant multiple.

5.5.2. Parameterized complexity of quantum knot invariants

Participant: Clément Maria.

In [53], we give a general fixed parameter tractable algorithm to compute quantum invariants of links presented by diagrams, whose complexity is singly exponential in the carving-width (or the tree-width) of the diagram. In particular, we get a $O(N^{3/2\mathrm{cw}}\mathrm{poly}(n))$ time algorithm to compute any Reshetikhin-Turaev invariant-derived from a simple Lie algebra g of a link presented by a planar diagram with n crossings and carving-width cw, and whose components are coloured with g-modules of dimension at most N. For example, this includes the Nth-coloured Jones polynomial and the Nth-coloured HOMFLYPT polynomial.

5.6. Miscellaneous

5.6.1. Material Coherence from Trajectories via Burau Eigenanalysis of Braids

Participant: David Cohen-Steiner.

In collaboration with Melissa Yeung and Mathieu Desbrun (Caltech).

In this paper [58], we provide a numerical tool to study material coherence from a set of 2D Lagrangian trajectories sampling a dynamical system, i.e., from the motion of passive tracers. We show that eigenvectors of the Burau representation of a topological braid derived from the trajectories have levelsets corresponding to components of the Nielsen-Thurston decomposition of the dynamical system. One can thus detect and identify clusters of space-time trajectories corresponding to coherent regions of the dynamical system by solving an eigenvalue problem. Unlike previous methods, the scalable computational complexity of our braid-based approach allows the analysis of large amounts of trajectories. Studying two-dimensional flows and their induced transport and mixing properties is key to geophysical studies of atmospheric and oceanic processes.

However, one often has only sparse tracer trajectories (e.g., positions of buoys in time) to infer the overall flow geometry. Fortunately, topological methods based on the theory of braid groups have recently been proposed to extract structures from such a sparse set of trajectories by measuring their entan-glement. This braid viewpoint offers sound foundations for the definition of coherent structures. Yet, there has been only limited efforts in developing practical tools that can leverage topological properties for the efficient analysis of flow structures: handling a larger number of tra-jectories remains computationally challenging. We contribute a new and simple computational tool to extract Lagrangian structures from sparse trajectories by noting that the eigenstructure of the Burau matrix representation of a braid of particle trajectories can be used to reveal coherent regions of the flows. Detection of clusters of space-time trajectories corresponding to coherent regions of the dynamical system can thus be achieved by solving a simple eigenvalue problem. This paper establishes the theoretical foundations behind this braid eigenanalysis approach, along with numerical validations on various flows.

5.6.2. Quantitative stability of optimal transport maps and linearization of the 2-Wasserstein space

Participants: Alex Delalande, Frédéric Chazal.

In collaboration with Quentin Mérigot (Institut de Mathématiques d'Orsay).

In [55], we study an explicit embedding of the set of probability measures into a Hilbert space, defined using optimal transport maps from a reference probability density. This embedding linearizes to some extent the 2-Wasserstein space, and enables the direct use of generic supervised and unsupervised learning algorithms on measure data. Our main result is that the embedding is (bi-)Holder continuous, when the reference density is uniform over a convex set, and can be equivalently phrased as a dimension-independent Hölder-stability results for optimal transport maps.

GAMBLE Project-Team

7. New Results

7.1. Non-Linear Computational Geometry

Participants: Laurent Dupont, Nuwan Herath Mudiyanselage, George Krait, Sylvain Lazard, Viviane Ledoux, Guillaume Moroz, Marc Pouget.

7.1.1. Clustering Complex Zeros of Triangular Systems of Polynomials

This work, presented at the CASC'19 Conference [23], gives the first algorithm for finding a set of natural ϵ-clusters of complex zeros of a regular triangular system of polynomials within a given polybox in $\mathbb{C}^n$, for any given $\epsilon > 0$. Our algorithm is based on a recent near-optimal algorithm of Becker et al (2016) for clustering the complex roots of a univariate polynomial where the coefficients are represented by number oracles. Our algorithm is based on recursive subdivision. It is local, numeric, certified and handles solutions with multiplicity. Our implementation is compared to well-known homotopy solvers on various triangular systems. Our solver always gives correct answers, is often faster than the homotopy solvers that often give correct answers, and sometimes faster than the ones that give sometimes correct results.

In collaboration with R. Imbach and C. Yap (Courant Institute of Mathematical Sciences, New York University, USA).

7.1.2. Numerical Algorithm for the Topology of Singular Plane Curves

We are interested in computing the topology of plane singular curves. For this, the singular points must be isolated. Numerical methods for isolating singular points are efficient but not certified in general. We are interested in developing certified numerical algorithms for isolating the singularities. In order to do so, we restrict our attention to the special case of plane curves that are projections of smooth curves in higher dimensions. In this setting, we show that the singularities can be encoded by a regular square system whose isolation can be certified by numerical methods. This type of curves appears naturally in robotics applications and scientific visualization. This work was presented at the EuroCG'19 Conference [24].

7.1.3. Reliable Computation of the Singularities of the Projection in $\mathbb{R}^3$ of a Generic Surface of $\mathbb{R}^4$

Computing efficiently the singularities of surfaces embedded in $\mathbb{R}^3$ is a difficult problem, and most state-of-the-art approaches only handle the case of surfaces defined by polynomial equations. Let F and G be C^∞ functions from $\mathbb{R}^4$ to $\mathbb{R}$ and $\mathcal{M} = \{(x, y, z, t) \in \mathbb{R}^4 \mid F(x, y, z, t) = G(x, y, z, t) = 0\}$ be the surface they define. Generically, the surface $\mathcal{M}$ is smooth and its projection Ω in $\mathbb{R}^3$ is singular. After describing the types of singularities that appear generically in Ω, we design a numerically well-posed system that encodes them. This can be used to return a set of boxes that enclose the singularities of Ω as tightly as required. As opposed to state-of-the art approaches, our approach is not restricted to polynomial mappings, and can handle trigonometric or exponential functions for example. This work was presented at the MACIS'19 Conference [19].

In collaboration with Sény Diatta (University Assane Seck of Ziguinchor, Senegal)

7.1.4. Evaluation of Chebyshev polynomials on intervals and application to root finding

In approximation theory, it is standard to approximate functions by polynomials expressed in the Chebyshev basis. Evaluating a polynomial f of degree n given in the Chebyshev basis can be done in $O(n)$ arithmetic operations using the Clenshaw algorithm. Unfortunately, the evaluation of f on an interval I using the Clenshaw algorithm with interval arithmetic returns an interval of width exponential in n. We describe a variant of the Clenshaw algorithm based on ball arithmetic that returns an interval of width quadratic in n for an interval of small enough width. As an application, our variant of the Clenshaw algorithm can be used to design an efficient root finding algorithm. This work was presented at the MACIS'19 Conference [21].

7.1.5. Using Maple to analyse parallel robots

We present the SIROPA Maple Library which has been designed to study serial and parallel manipulators at the conception level. We show how modern algorithms in Computer Algebra can be used to study the workspace, the joint space but also the existence of some physical capabilities w.r.t. to some design parameters left as degree of freedom for the designer of the robot. This work was presented at the Maple Conference 2019 [18].

In collaboration with Philippe Wenger, Damien Chablat (Laboratoire des Sciences du Numérique de Nantes, UMR CNRS 6004) and Fabrice Rouillier (project team OURAGAN *)*

7.2. Non-Euclidean Computational Geometry

Participants: Vincent Despré, Yan Garito, Elies Harington, Benedikt Kolbe, Georg Osang, Monique Teillaud, Gert Vegter.

7.2.1. Flipping Geometric Triangulations on Hyperbolic Surfaces

We consider geometric triangulations of surfaces, i.e., triangulations whose edges can be realized by disjoint locally geodesic segments. We prove that the flip graph of geometric triangulations with fixed vertices of a flat torus or a closed hyperbolic surface is connected. We give upper bounds on the number of edge flips that are necessary to transform any geometric triangulation on such a surface into a Delaunay triangulation [28].

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

7.2.2. Computing the Geometric Intersection Number of Curves

The geometric intersection number of a curve on a surface is the minimal number of self-intersections of any homotopic curve, i.e. of any curve obtained by continuous deformation. Given a curve c represented by a closed walk of length at most ℓ on a combinatorial surface of complexity n we describe simple algorithms to compute the geometric intersection number of c in $O(n+\ell^2)$ time, construct a curve homotopic to c that realizes this geometric intersection number in $O(n+\ell^4)$ time, decide if the geometric intersection number of c is zero, i.e. if c is homotopic to a simple curve, in $O(n+\ell\log(\ell))$ time [14].

In collaboration with Francis Lazarus (University of Grenoble).

7.3. Probabilistic Analysis of Geometric Data Structures and Algorithms

Participants: Olivier Devillers, Charles Duménil, Xavier Goaoc, Fernand Kuiebove Pefireko, Ji Won Park.

7.3.1. Expected Complexity of Routing in Θ6 and Half-Θ6 Graphs

We study online routing algorithms on the Θ6-graph and the half-Θ6-graph (which is equivalent to a variant of the Delaunay triangulation). Given a source vertex s and a target vertex t in the Θ6-graph (resp. half-Θ6-graph), there exists a deterministic online routing algorithm that finds a path from s to t whose length is at most 2 st (resp. 2.89 st) which is optimal in the worst case [Bose et al., SIAM J. on Computing, 44(6)]. We propose alternative, slightly simpler routing algorithms that are optimal in the worst case and for which we provide an analysis of the average routing ratio for the Θ6-graph and half-Θ6-graph defined on a Poisson point process. For the Θ6-graph, our online routing algorithm has an expected routing ratio of 1.161 (when s and t random) and a maximum expected routing ratio of 1.22 (maximum for fixed s and t where all other points are random), much better than the worst-case routing ratio of 2. For the half-Θ6-graph, our memoryless online routing algorithm has an expected routing ratio of 1.43 and a maximum expected routing ratio of 1.58. Our online routing algorithm that uses a constant amount of additional memory has an expected routing ratio of 1.34 and a maximum expected routing ratio of 1.40. The additional memory is only used to remember the coordinates of the starting point of the route. Both of these algorithms have an expected routing ratio that is much better than their worst-case routing ratio of 2.89 [27].

In collaboration with Prosenjit Bose (University Carleton) and JeanLou De Carufel (University of Ottawa)

7.3.2. A Poisson sample of a smooth surface is a good sample

The complexity of the 3D-Delaunay triangulation (tetrahedralization) of n points distributed on a surface ranges from linear to quadratic. When the points are a deterministic good sample of a smooth compact generic surface, the size of the Delaunay triangulation is $O(n \log n)$. Using this result, we prove that when points are Poisson distributed on a surface under the same hypothesis, whose expected number of vertices is λ, the expected size is $O(\lambda \log_2 \lambda)$ [22].

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. 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 [29].

In collaboration with Philippe Duchon (Université de Bordeaux) and Marc Glisse (project team DATASHAPE *).*

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

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 on polyhedral surfaces in $\mathbb{E}^3$ has linear complexity, as opposed to a worst-case quadratic complexity. The standard analysis does not provide accurate bounds on the complexity of such cases and we aim at establishing such bounds. More precisely, we will show that, in the case of nicely distributed points on polyhedral surfaces, 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. Our proofs also work for some other notions of nicely distributed point sets, such as (ϵ, κ)-samples. Along the way, we prove a probabilistic lemma for sampling without replacement, which may be of independent interest [16], [26].

In collaboration with Jean-Daniel Boissonnat, Kunal Dutta and Marc Glisse (project team DATASHAPE *).*

7.3.5. Random polytopes and the wet part for arbitrary probability distributions

We examine how the measure and the number of vertices of the convex hull of a random sample of n points from an arbitrary probability measure in $\mathbb{R}^d$ relates to the wet part of that measure. This extends classical results for the uniform distribution from a convex set [Bárány and Larman 1988]. The lower bound of Bárány and Larman continues to hold in the general setting, but the upper bound must be relaxed by a factor of $\log n$. We show by an example that this is tight [25].

In collaboration with Imre Barany (Rényi Institute of Mathematics) Matthieu Fradelizi (Laboratoire d'Analyse et de Mathématiques Appliquées) Alfredo Hubard (Laboratoire d'Informatique Gaspard-Monge) Günter Rote (Institut für Informatik, Berlin)

7.4. Discrete Geometric structures

Participants: Xavier Goaoc, Galatée Hemery Vaglica.

7.4.1. Shatter functions with polynomial growth rates

We study how a single value of the shatter function of a set system restricts its asymptotic growth. Along the way, we refute a conjecture of Bondy and Hajnal which generalizes Sauer's Lemma. [12]

7.4.2. The discrete yet ubiquitous theorems of Caratheodory, Helly, Sperner, Tucker, and Tverberg

We discuss five discrete results: the lemmas of Sperner and Tucker from combinatorial topology and the theorems of Carathéodory, Helly, and Tverberg from combinatorial geometry. We explore their connections and emphasize their broad impact in application areas such as game theory, graph theory, mathematical optimization, computational geometry, etc. [13]

7.4.3. Shellability is NP-complete

We prove that for every $d \geq 2$, deciding if a pure, d-dimensional, simplicial complex is shellable is NP-hard, hence NP-complete. This resolves a question raised, e.g., by Danaraj and Klee in 1978. Our reduction also yields that for every $d \geq 2$ and $k \geq 0$, deciding if a pure, d-dimensional, simplicial complex is k-decomposable is NP-hard. For $d \geq 3$, both problems remain NP-hard when restricted to contractible pure d-dimensional complexes. Another simple corollary of our result is that it is NP-hard to decide whether a given poset is CL-shellable. [15]

7.4.4. An Experimental Study of Forbidden Patterns in Geometric Permutations by Combinatorial Lifting

We study the problem of deciding if a given triple of permutations can be realized as geometric permutations of disjoint convex sets in $\mathbb{R}^3$. We show that this question, which is equivalent to deciding the emptiness of certain semi-algebraic sets bounded by cubic polynomials, can be "lifted" to a purely combinatorial problem. We propose an effective algorithm for that problem, and use it to gain new insights into the structure of geometric permutations. [20]

7.5. Classical Computational Geometry

Participants: Olivier Devillers, Sylvain Lazard, Leo Valque.

7.5.1. Rounding Meshes

Let $\mathcal{P}$ be a set of n polygons in $\mathbb{R}^3$, each of constant complexity and with pairwise disjoint interiors. We previously proposed [5] a rounding algorithm that maps $\mathcal{P}$ to a simplicial complex $\mathcal{Q}$ whose vertices have integer coordinates such that 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 built 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. We developed [30] the first implementation of this algorithm, which is also the first implementation for rounding a mesh on a grid (whose fineness is independent of the input size) while preserving reasonable geometric and topological properties. We also provided some insight that this algorithm and implementation have practical average complexity in $O(n\sqrt{n})$ on “real data", which has to be compared to its $O(n^{15})$ worst-case time complexity. Our implementation is still too slow to be used in practice but it provides a good proof of concept.

7.5.2. Hardness results on Voronoi, Laguerre and Apollonius diagrams

We show that converting Apollonius and Laguerre diagrams from an already built Voronoi diagram of a set of n points in 2D requires at least $\Omega(n \log n)$ computation time. We also show that converting an Apollonius diagram of a set of n weighted points in 2D from a Laguerre diagram and vice-versa requires at least $\Omega(n \log n)$ computation time as well. Furthermore , we present a very simple randomized incremental construction algorithm that takes expected $O(n \log n)$ computation time to build an Apollonius diagram of non-overlapping circles in 2D [17].

In collaboration with Kevin Buchin (TU Eindhoven), Pedro de Castro (University Pernanbuco), and Menelaos Karavelas (University Heraklion).

GRACE Project-Team

6. New Results

6.1. Error Locating pairs

Participants: Alain Couvreur, Isabella Panaccione.

Algebraic codes such as Reed–Solomon codes and algebraic geometry codes benefit from efficient decoding algorithms permitting to correct errors up to half the minimum distance and sometimes beyond. In 1992, Pellikaan proved that many **unique** decoding could be unified using an object called *Error correcting pair*. In short, given an error correcting code $\mathcal{C}$, an error correcting pair for $\mathcal{C}$ is a pair of codes $(\mathcal{A}, \mathcal{B})$ whose component wise product $\mathcal{A} * \mathcal{B}$ is contained in the dual code $\mathcal{C}^{\perp}$ and such that $\mathcal{A}, \mathcal{B}$ satisfy some constraints of dimension and minimum distance.

On the other hand, in the late 90's, after the breakthrough of Sudan and Guruswami Sudan the question of list decoding permitting to decode beyond half the minimum distance. In a recently submitted article, A. Couvreur and I. Panaccione [15] proposed a unified point of view for probabilistic decoding algorithms decoding beyond half the minimum distance. Similarly to Pellikaan's result, this framework applies to any code benefiting from an *error locating pair* which is a relaxed version of error correcting pairs.

6.2. Factoring oracles

Participants: François Morain, Benjamin Smith, Guénaël Renault.

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 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, The manuscript [18] is currently being refereed.

LFANT Project-Team

6. New Results

6.1. Cryptographic Protocols

Participants: Guilhem Castagnos, Ida Tucker.

In [20], G. Castagnos, D. Catalano, F. Laguillaumie, F. Savasta and I. Tucker propose a new cryptographic protocol to compute ECDSA signatures with two parties.

ECDSA (Elliptic Curves Digital Signature Algorithm) is a widely adopted standard for electronic signatures. For instance, it is used in the TLS (Transport Layer Security) protocol and in many cryptocurrencies such as Bitcoin. For cryptocurrencies, ECDSA is used in order to sign the transactions: if Alice wants to give n bitcoins to Bob, she uses her secret key to sign with ECDSA a bit string encoding this information.

As a result, if the secret key of Alice is stolen, for example if her computer is compromised, an attacker can stole all her bitcoins. A common solution to this problem is to share the key on multiple devices, for example a laptop and a mobile phone. Both devices must collaborate in order to issue a signature, and if only one device is compromised, no information on the key is leaked. This setting belongs to the area of secure multiparty computation.

There have been recent proposals to construct 2 party variants of ECDSA signatures but constructing efficient protocols proved to be much harder than for other signature schemes. The main reason comes from the fact that the ECDSA signing protocol involves a complex equation compared to other signatures schemes. Lindell recently managed to get an efficient solution using the linearly homomorphic cryptosystem of Paillier. However his solution has some drawbacks, for example the security proof resorts to a non-standard interactive assumption.

By using another approach based on hash proofs systems we obtain a proof that relies on standard assumptions. Moving to concrete constructions, we show how to instantiate our framework using class groups of imaginary quadratic fields. Our implementations show that the practical impact of dropping such interactive assumptions is minimal. Indeed, while for 128-bit security our scheme is marginally slower than Lindell's, for 256-bit security it turns out to be better both in key generation and signing time. Moreover, in terms of communication cost, our implementation significantly reduces both the number of rounds and the transmitted bits without exception.

This paper was presented at the CRYPTO Conference 2019, and is part of the ALAMBIC project.

6.2. Coding Theory

Participants: Xavier Caruso, Aurel Page.

In [29], Xavier Caruso developed a theory of residues for skew rational functions (which are, by definition, the quotients of two skew polynomials), proving in particular a skew analogue of the residue formula and a skew analogue of the classical formula of change of variables for residues. He then used his theory to define and study a linearized version of Goppa codes. He showed that these codes meet the Singleton bound (for the sum-rank metric) and are the duals of the linearized Reed–Solomon codes defined recently by Martínez-Peñas. Efficient encoding and decoding algorithms are also designed.

C. Maire and A. Page updated the preprint *Error-correcting codes based on non-commutative algebras* [33] according to the comments of referees.

6.3. Number fields

Participants: Razvan Barbulescu, Jean-Marc Couveignes, Jean-Paul Cerri, Pierre Lezowski.

In [30], Jean-Marc Couveignes constructs small models of number fields and deduces a better bound for the number of number fields of given degree n and discriminant bounded by H. This work improves on previous results by Schmidt and Ellenberg-Venkatesh. Schmidt obtains a bound $H^{\frac{n+2}{4}}$ times a function of n. Ellenberg and Venkatesh obtain a bound $H^{\exp(O(\sqrt{\log n}))}$ times a function of n. The new idea is to combine geometry of numbers and interpolation theory to produces small projective models and lower the exponent of H down to $O(\log^3 n)$. A key point is to look for local equations rather than a full set of generators of the ideal of these models.

In [12], Razvan Barbulescu in a joint work with Jishnu Ray (University of British Columbia, Vancouver) brings elements to support Greenberg's p-rationality conjecture. On the theoretical side, they propose a new family proven to be p-rational. On the algorithmic side, the compare the tools to enumerate number fields of given abelian Galois group and of computing class numbers, and extend the experiments on the Cohen-Lenstra-Martinet conjectures.

In collaboration with Pierre Lezowski, Jean-Paul Cerri has studied in [15] 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. Additionally, 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 non-commutative 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.

6.4. Modular forms and L-functions

Participant: Henri Cohen.

Members of the team have taken part in an international autumn school on computational number theory at the Izmir Institute of Technology (IZTECH) in 2017. Henri Cohen has transformed his two lectures in book chapters. The text on modular forms [23] presents the (of course extremely condensed) view of the book [6] he has coauthored. The chapter on L-functions [24] is closely related to new developments in PARI/GP.

In [25] the same author explains how to compute Fourier expansions at all cusps of any modular form of integral or half-integral weight thanks to a theorem of Borisov–Gunnells and explicit expansions of Eisenstein series at all cusps. Using this, he gives a number of methods for computing arbitrary Petersson products. Implementations in our PARI/GP software are also described.

A complementary approach using modular symbols is used in [14] by Karim Belabas, Dominique Bernardi and Bernadette Perrin-Riou to compute Manin's constant and the modular degree of elliptic curves defined over $\mathbb{Q}$.

6.5. p-adic rings and geometry

Participant: Xavier Caruso.

In [19], Xavier Caruso, Tristan Vaccon and Thibaut Verron laid the foundations of an algorithmic treatment of rigid p-adic geometry by introducting and studing Gröbner bases over Tate algebras. In addition, they designed a Buchberger-like and a F4-like algorithm for computing such Gröbner bases.

In [22], Xavier Caruso presents a survey on Fontaine's theory of p-adic period rings. These notes are based on a course given jointly by Laurent Berger and Xavier Caruso in Rennes in 2014; their aim is to detail the construction of the rings B_{crys} and B_{dR} (and some of their variants) and state several comparison theorems between étale and crystalline or de Rham cohomologies for p-adic algebraic varieties.

6.6. Geometry

Participant: Aurel Page.

The paper [13], *Can you hear the homology of 3-dimensional drums?* by A. Bartel and A. Page was published in Commentarii Mathematici Helvetici.

6.7. Complex multiplication of abelian varieties and elliptic curves

Participants: Razvan Barbulescu, Sorina Ionica, Chloe Martindale, Enea Milio, Damien Robert.

In [16], Sorina Ionica, former postdoc of the team, and Emmanuel Thomé look at the structure of isogeny graphs of genus 2 Jacobians with maximal real multiplication. They generalise a result of Kohel's describing the structure of the endomorphism rings of the isogeny graph of elliptic curves. Their setting considers genus 2 jacobians with complex multiplication, with the assumptions that the real multiplication subring is maximal and has class number 1. Over finite fields, they derive a depth first search algorithm for computing endomorphism rings locally at prime numbers, if the real multiplication is maximal.

Antonin Riffaut examines in [18] whether there are relations defined over $\mathbb{Q}$ that link (additively or multiplicatively) different singular moduli $j(\tau)$, invariants of elliptic curves with complex multiplication by different quadratic rings.

In [34], Chloe Martindale presents an algorithm to compute higher dimensional Hilbert modular polynomials. She also explains applications of this algorithm to point counting, walking on isogeny graphs, and computing class polynomials.

In [28], Razvan Barbulescu and Sudarshan Shinde (Sorbonne Université) make a complete list of the 1525 infinite families of elliptic curves without CM which have a particular behaviour in the ECM factoring algorithm, the 20 previously known families having been found by ad-hoc methods. The new idea was to use the characterisation of ECM-friendly families in terms of their Galois image and to use the recent progress in the topic of Mazur's program. In particular, for some of the families mentioned theoretical in the literature the article offers the first publication of explicite equations.

E. Milio and D. Robert updated their paper [35] on computing cyclic modular polynomials.

6.8. Pairings

Participant: Razvan Barbulescu.

In [27], Razvan Barbulescu in a joint work with Nadia El Mrabet (École des Mines de Saint-Étienne) et Loubna Ghammam (Bosch) makes a review of the families of elliptic curves for pairing-based cryptology. This was necessary after the invention of a new variant of the NFS algorithm in 2016 by Barbulescu and Taechan Kim, which showed that the previously used key sizes for pairings were insecure. The novelty of this review article is double : first they consider a large number of families, some of which were never analysed in the literature because they were not likely to be the best and secondly they combine in the same article the security analysis of each family with a non-optimized implementation. This allows the industry to select a different family for each type of utilisation of pairings.

6.9. Multiprecision arithmetic

Participant: Fredrik Johansson.

In [17], 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.

In [26], 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.

In [21], Fredrik Johansson develops algorithms for real and complex dot product and matrix multiplication in arbitrary-precision floating-point and ball arithmetic. The new methods are implemented in Arb and significantly speed up polynomial operations and linear algebra in high precision.

OURAGAN Project-Team

7. New Results

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

In [18], we present new computer algebra based methods for testing the structural stability of n-D discrete linear systems (with n at least 2). More precisely, we 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.

7.2. Computing period matrices and the Abel-Jacobi map of superelliptic curves

In [24], we present an algorithm for the computation of period matrices and the Abel-Jacobi map of complex superelliptic curves given by an equation y m = f (x). It relies on rigorous numerical integration of differentials between Weierstrass points, which is done using Gauss method if the curve is hyperelliptic (m = 2) or the Double-Exponential method. The algorithm is implemented and makes it possible to reach thousands of digits accuracy even on large genus curves.

7.3. Voronoi diagram of orthogonal polyhedra in two and three dimensions

Voronoi diagrams are a fundamental geometric data structure for obtaining proximity relations. In [28], we consider collections of axis-aligned orthogonal polyhedra in two and three-dimensional space under the max-norm, which is a particularly useful scenario in certain application domains. We construct the exact Voronoi diagram inside an orthogonal polyhedron with holes defined by such polyhedra. Our approach avoids creating full-dimensional elements on the Voronoi diagram and yields a skeletal representation of the input object. We introduce a complete algorithm in 2D and 3D that follows the subdivision paradigm relying on a bounding-volume hierarchy; this is an original approach to the problem. The complexity is adaptive and comparable to that of previous methods. Under a mild assumption it is $O(n/\Delta)$ in 2D or $O(n.\alpha^2/\Delta^2)$ in 3D, where n is the number of sites, namely edges or facets resp., Δ is the maximum cell size for the subdivision to stop, and α bounds vertex cardinality per facet. We also provide a numerically stable, open-source implementation in Julia, illustrating the practical nature of our algorithm.

7.4. A symbolic computation approach towards the asymptotic stability analysis of differential systems with commensurate delays

In [30], the work aims at studying the asymptotic stability of retarded type linear differential systems with commensurate delays. Within the frequency-domain approach, it is well-known that the asymptotic stability of such a system is ensured by the condition that all the roots of the corresponding quasipolynomial have negative real parts. A classical approach for checking this condition consists in computing the set of critical zeros of the quasipolynomial, i.e., the roots (and the corresponding delays) of the quasipolynomial that lie on the imaginary axis, and then analyzing the variation of these roots with respect to the variation of the delay. Following this approach, based on solving algebraic systems techniques, we propose a certified and efficient symbolic-numeric algorithm for computing the set of critical roots of a quasipolynomial. Moreover, using recent algorithmic results developed by the computer algebra community, we present an efficient algorithm for the computation of Puiseux series at a critical zero which allows us to finely analyze the stability of the system with respect to the variation of the delay. Explicit examples are given to illustrate our algorithms.

7.5. On the computation of stabilizing controllers of multidimensional systems

In [25], we consider the open problem consisting in the computation of stabilizing controllers of an internally stabilizable MIMO multidimensional system. Based on homological algebra and the so-called *Polydisk Nullstellensatz*, we propose a general method towards the explicit computation of stabilizing controllers. We show how the homological algebra methods over the ring of structurally stable SISO multidimensional transfer functions can be made algorithmic based on standard Gröbner basis techniques over polynomial rings. The problem of computing stabilizing controllers is then reduced to the problem of obtaining an effective version of the Polydisk Nullstellensatz which, apart from a few cases, stays open and will be studied in forthcoming publications.

7.6. Algebraic aspects of the exact signal demodulation problem

In [29], we introduce a general class of problems originating from gearbox vibration analysis. Based on a previous work where demodulation was formulated as a matrix approximation problem, we study the specific case applicable to amplitude and phase demodulation. This problem can be rewritten as a polynomial system. Based on algebraic methods such as linear algebra and homological algebra, we focus on the characterization of the problem and solve it in the noise-free case.

7.7. General closed-form solutions of the position self-calibration problem

The work in [36] investigates the anchors and sources position self-calibration problem in the 3D space based on range measurements and without any prior restriction on the network configuration. Using a well known low-rank property of Euclidean distance matrices, we first reduce the problem to finding 12 unknowns ascribed in a 3×3 transformation matrix and a 3×1 translation vector. In order to estimate them, we then introduce a polynomial parametrization with 9 unknowns that are estimated by solving a linear system. Afterwards, we identify an intrinsic matrix polynomial system that encodes the solution set of the problem and provide a direct method for solving it. The resulting procedure is simple and straightforward to implement using standard numerical tools. We also show that closed-form solutions can always be obtained when the reference frame is fixed. This is illustrated by adopting reference frames from the literature and by introducing a triangular reference frame whose constraints are imposed only on one position set (anchor or source). Experimental results on synthetic and real sound data show that the proposed closed-form solutions efficiently solve the position self-calibration problem.

7.8. Certified lattice reduction

Quadratic form reduction and lattice reduction are fundamental tools in computational number theory and in computer science, especially in cryptography. The celebrated Lenstra–Lenstra–Lovász reduction algorithm (so-called LLL) has been improved in many ways through the past decades and remains one of the central methods used for reducing integral lattice basis. In particular, its floating-point variants—where the rational arithmetic required by Gram–Schmidt orthogonalization is replaced by floating-point arithmetic—are now the fastest known. However, the systematic study of the reduction theory of real quadratic forms or, more generally, of real lattices is not widely represented in the literature. When the problem arises, the lattice is usually replaced by an integral approximation of (a multiple of) the original lattice, which is then reduced. While practically useful and proven in some special cases, this method doesn't offer any guarantee of success in general. In [22], we present an adaptive-precision version of a generalized LLL algorithm that covers this case in all generality. In particular, we replace floating-point arithmetic by Interval Arithmetic to certify the behavior of the algorithm. We conclude by giving a typical application of the result in algebraic number theory for the reduction of ideal lattices in number fields.

7.9. Using Maple to analyse parallel robots

In [27], we present the SIROPA Maple Library which has been designed to study serial and parallel manipulators at the conception level. We show how modern algorithms in Computer Algebra can be used to study the workspace, the joint space but also the existence of some physical capabilities w.r.t. to some design parameters left as degree of freedom for the designer of the robot.

7.10. On the effective computation of stabilizing controllers of 2D systems

In [26], we show how stabilizing controllers for 2D systems can effectively be computed based on computer algebra methods dedicated to polynomial systems, module theory and homological algebra. The complete chain of algorithms for the computation of stabilizing controllers, implemented in Maple, is illustrated with an explicit example.

7.11. Updating key size estimations for pairings

Recent progress on NFS imposed a new estimation of the security of pairings. In [15], 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.

7.12. Matrix formulae for Resultants and Discriminants of Bivariate Tensor-product Polynomials

The construction of optimal resultant formulae for polynomial systems is one of the main areas of research in computational algebraic geometry. However, most of the constructions are restricted to formulae for unmixed polynomial systems, that is, systems of polynomials which all have the same support. Such a condition is restrictive, since mixed systems of equations arise frequently in many problems. Nevertheless, resultant formulae for mixed polynomial systems is a very challenging problem. In [19], we present a square, Koszul-type, matrix, the determinant of which is the resultant of an arbitrary (mixed) bivariate tensor-product polynomial system. The formula generalizes the classical Sylvester matrix of two univariate polynomials, since it expresses a map of degree one, that is, the elements of the corresponding matrix are up to sign the coefficients of the input polynomials. Interestingly, the matrix expresses a primal-dual multiplication map, that is, the tensor product of a univariate multiplication map with a map expressing derivation in a dual space. In addition we prove an impossibility result which states that for tensor-product systems with more than two (affine) variables there are no universal degree-one formulae, unless the system is unmixed. Last but not least, we present applications of the new construction in the efficient computation of discriminants and mixed discriminants.

7.13. Separation bounds for polynomial systems

In [21], we rely on aggregate separation bounds for univariate polynomials to introduce novel worst-case separation bounds for the isolated roots of zero-dimensional, positive-dimensional, and overde- termined polynomial systems. We exploit the structure of the given system, as well as bounds on the height of the sparse (or toric) resultant, by means of mixed volume, thus establishing adaptive bounds. Our bounds improve upon Canny's Gap theorem [9]. Moreover, they exploit sparseness and they apply without any assumptions on the input polynomial system. To evaluate the quality of the bounds, we present polynomial systems whose root separation is asymptotically not far from our bounds. We apply our bounds to three problems. First, we use them to estimate the bitsize of the eigenvalues and eigenvectors of an integer matrix; thus we provide a new proof that the problem has polynomial bit complexity. Second, we bound the value of a positive polynomial over the simplex: we improve by at least one order of magnitude upon all existing bounds. Finally, we asymptotically bound the number of steps of any purely subdivision-based algorithm that isolates all real roots of a polynomial system.

7.14. On the maximal number of real embeddings of minimally rigid graphs in $\mathbb{R}^2$, $\mathbb{R}^3$ and S^2

Rigidity theory studies the properties of graphs that can have rigid embeddings in a euclidean space $\mathbb{R}^d$ or on a sphere and other manifolds which in addition satisfy certain edge length constraints. One of the major open problems in this field is to determine lower and upper bounds on the number of realizations with respect to a given number of vertices. This problem is closely related to the classification of rigid graphs according to their maximal number of real embeddings. In [17], we are interested in finding edge lengths that can maximize the number of real embeddings of minimally rigid graphs in the plane, space, and on the sphere. We use algebraic formulations to provide upper bounds. To find values of the parameters that lead to graphs with a large number of real realizations, possibly attaining the (algebraic) upper bounds, we use some standard heuristics and we also develop a new method inspired by coupler curves. We apply this new method to obtain embeddings in $\mathbb{R}^3$. One of its main novelties is that it allows us to sample efficiently from a larger number of parameters by selecting only a subset of them at each iteration. Our results include a full classification of the 7-vertex graphs according to their maximal numbers of real embeddings in the cases of the embeddings in $\mathbb{R}^2$ and $\mathbb{R}^3$, while in the case of S^2 we achieve this classification for all 6-vertex graphs. Additionally, by increasing the number of embeddings of selected graphs, we improve the previously known asymptotic lower bound on the maximum number of realizations.

7.15. Multilinear Polynomial Systems: Root Isolation and Bit Complexity

In [20], we exploit structure in polynomial system solving by considering polyno-mials that are linear in subsets of the variables. We focus on algorithms and their Boolean complexity for computing isolating hyperboxes for all the isolated complex roots of well-constrained, unmixed systems of multilinear polynomials based on resultant methods. We enumerate all expressions of the multihomogeneous (or multigraded) resultant of such systems as a determinant of Sylvester-like matrices, aka generalized Sylvester matrices. We construct these matrices by means of Weyman homological complexes, which generalize the Cayley-Koszul complex. The computation of the determinant of the resultant matrix is the bottleneck for the overall complexity. We exploit the quasi-Toeplitz structure to reduce the problem to efficient matrix-vector multiplication, which corresponds to multivariate polynomial multiplication, by extending the seminal work on Macaulay matrices of Canny, Kaltofen, and Yagati [9] to the multi-homogeneous case. We compute a rational univariate representation of the roots, based on the primitive element method. In the case of 0-dimensional systems we present a Monte Carlo algorithm with probability of success $1 - 1/2^{\eta}$, for a given $\eta \geq 1$, and bit complexity $O_B(n^2 D^{4+\epsilon}(n^{N+1} + \tau) + nD^{2+\epsilon}\eta(D + \eta))$ for any $\epsilon > 0$, where n is the number of variables, D equals the multilinear Bézout bound, N is the number of variable subsets, and τ is the maximum coefficient bitsize. We present an algorithmic variant to compute the isolated roots of overdetermined and positive-dimensional systems. Thus our algorithms and complexity analysis apply in general with no assumptions on the input.

POLSYS Project-Team

6. New Results

6.1. Fundamental algorithms and structured polynomial systems

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 [3], our first contribution is to make the Berlekamp–Massey–Sakata algorithm more efficient by making it adaptive to avoid some useless relation test-ings. 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.

The problem of finding $m \times s$ matrices (with $m \geq s$) of rank r in a real affine subspace of dimension n has many applications in information and systems theory, where low rank is synonymous of structure and parsimony. In [8], we design computer algebra algorithms to solve this problem efficiently and exactly: the input are the rational coefficients of the matrices spanning the affine subspace as well as the expected maximum rank, and the output is a rational parametrization encoding a finite set of points that intersects each connected component of the low rank real algebraic set. The complexity of our algorithm is studied thoroughly. It is essentially polynomial in $n + m(s - r)$; it improves on the state-of-the-art in the field. Moreover, computer experiments show the practical efficiency of our approach.

Gröbner bases is one the most powerful tools in algorithmic non-linear algebra. Their computation is an intrinsically hard problem with a complexity at least single exponential in the number of variables. However, in most of the cases, the polynomial systems coming from applications have some kind of structure. For example , several problems in computer-aided design, robotics, vision, biology , kinematics, cryptography, and optimization involve sparse systems where the input polynomials have a few non-zero terms. In [16], our approach to exploit sparsity is to embed the systems in a semigroup algebra and to compute Gröbner bases over this algebra. Up to now, the algorithms that follow this approach benefit from the sparsity only in the case where all the polynomials have the same sparsity structure, that is the same Newton polytope. We introduce the first algorithm that overcomes this restriction. Under regularity assumptions, it performs no redundant computations. Further, we extend this algorithm to compute Gröbner basis in the standard algebra and solve sparse polynomials systems over the torus $(\mathbb{C}^{\star})^{n}$. The complexity of the algorithm depends on the Newton polytopes.

In [10], we consider the problem of approximating numerically the moments and the supports of measures which are invariant with respect to the dynamics of continuous- and discrete-time polynomial systems, under semialgebraic set constraints. First, we address the problem of approximating the density and hence the support of an invariant measure which is absolutely continuous with respect to the Lebesgue measure. Then, we focus on the approximation of the support of an invariant measure which is singular with respect to the Lebesgue measure. Each problem is handled through an appropriate reformulation into a linear optimization problem over measures, solved in practice with two hierarchies of finite-dimensional semidefinite moment-sum-of-square relaxations, also called Lasserre hierarchies. Under specific assumptions, the first Lasserre hierarchy allows to approximate the moments of an absolutely continuous invariant measure as close as desired and to extract a sequence of polynomials converging weakly to the density of this measure. The second Lasserre hierarchy allows to approximate as close as desired in the Hausdorff metric the support of a singular invariant measure with the level sets of the Christoffel polynomials associated to the moment matrices of this measure. We also present some application examples together with numerical results for several dynamical systems admitting either absolutely continuous or singular invariant measures.

6.2. Solving systems over the reals and applications

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 [11], 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.

[9] describes our freely distributed Maple library SPECTRA, for Semidefinite Programming solved Exactly with Computational Tools of Real Algebra. It solves linear matrix inequalities with symbolic computation in exact arithmetic and it is targeted to small-size, possibly degenerate problems for which symbolic infeasibility or feasibility certificates are required.

Let $S \subset \mathbb{R}^n$ be a compact basic semi-algebraic set defined as the real solution set of multivariate polynomial inequalities with rational coefficients. In [19], we design an algorithm which takes as input a polynomial system defining S and an integer $p \geq 0$ and returns the n-dimensional volume of S at absolute precision 2^{-p}. Our algorithm relies on the relationship between volumes of semi-algebraic sets and periods of rational integrals. It makes use of algorithms computing the Picard-Fuchs differential equation of appropriate periods, properties of critical points, and high-precision numerical integration of differential equations. The algorithm runs in essentially linear time with respect to p. This improves upon the previous exponential bounds obtained by Monte-Carlo or moment-based methods. Assuming a conjecture of Dimca, the arithmetic cost of the algebraic subroutines for computing Picard-Fuchs equations and critical points is singly exponential in n and polynomial in the maximum degree of the input.

Let $\mathbf{f} = (f_1, ..., f_s)$ be a sequence of polynomials in $\mathbb{Q}[X_1, ..., X_n]$ of maximal degree D and $V \subset \mathbb{C}^n$ be the algebraic set defined by $\mathbf{f}$ and r be its dimension. The real radical $\sqrt{\langle \mathbf{f} \rangle}$ associated to $\mathbf{f}$ is the largest ideal which defines the real trace of V. When V is smooth, we show in [13], that $\sqrt[re]{\langle \mathbf{f} \rangle}$, has a finite set of generators with degrees bounded by $\deg V$. Moreover, we present a probabilistic algorithm of complexity $(snD^n)^{O(1)}$ to compute the minimal primes of $\sqrt[re]{\langle \mathbf{f} \rangle}$. 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 $V \cap \mathbb{R}^n$.

Let $(g_1, ..., g_p)$ in $\mathbb{Q}[X_1, ..., X_n]$ and S be the basic closed semi-algebraic set defined by $g_1 \geq 0, ..., g_p \geq 0$. The S-radical of $\langle \mathbf{f} \rangle$, which is denoted by $\sqrt[S]{\langle \mathbf{f} \rangle}$, is the ideal associated to the Zariski closure of $V \cap S$.

We give a probabilistic algorithm to compute rational parametrizations of all irreducible components of that Zariski closure, hence encoding $\sqrt[\mathbb{R}]{\langle \mathbf{f} \rangle}$. Assuming now that D is the maximum of the degrees of the f_i's and the g_i's, this algorithm runs in time $2^p(s+p)^{O(1)}(nD)^{O(rn2^r)}$.

Experiments are performed to illustrate and show the efficiency of our approaches on computing real radicals.

In [14], we consider the second-order discontinuous differential equation $y'' + \eta \mathrm{sgn}(y) = \theta y + \alpha \sin(\beta t)$ where the parameters $\eta, \theta, \alpha, \beta$ are real. The main goal is to discuss the existence of periodic solutions. Under explicit conditions, the number of such solutions is given. Furthermore, for each of these periodic solutions, an explicit formula is provided.

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

6.3.1. Algebraic Cryptanalysis of a Quantum Money Scheme – The Noisy Case.

At STOC 2012, Aaronson and Christiano proposed a noisy and a noiseless version of the first public-key quantum money scheme endowed with a security proof. [5] addresses the so-called noisy hidden subspaces problem, on which the noisy version of their scheme is based. The first contribution of this work is a non-quantum cryptanalysis of the above-mentioned noisy quantum money scheme extended to prime fields $\mathbb{F}$, with $|\mathbb{F}| \neq 2$, that runs in randomised polynomial time. This finding is supported with experimental results showing that, in practice, the algorithm presented is efficient and succeeds with overwhelming probability. The second contribution is a non-quantum randomised polynomial-time cryptanalysis of the noisy quantum money scheme over $\mathbb{F}_2$ succeeding with a certain probability for values of the noise lying within a certain range. This result disproves a conjecture made by Aaronson and Christiano about the non-existence of an algorithm that solves the noisy hidden subspaces problem over $\mathbb{F}_2$ and succeeds with such probability.

6.3.2. On the Complexity of MQ in the Quantum Setting.

In August 2015 the cryptographic world was shaken by a sudden and surprising announcement by the US National Security Agency NSA concerning plans to transition to post-quantum algorithms. Since this announcement post-quantum cryptography has become a topic of primary interest for several standardization bodies. The transition from the currently deployed public-key algorithms to post-quantum algorithms has been found to be challenging in many aspects. In particular the problem of evaluating the quantum-bit security of such post-quantum cryptosystems remains vastly open. Of course this question is of primarily concern in the process of standardizing the post-quantum cryptosystems. In [21] we consider the quantum security of the problem of solving a system of m *Boolean multivariate quadratic equations in* n *variables* (MQb); a central problem in post-quantum cryptography. When $n = m$, under a natural algebraic assumption, we present a Las-Vegas quantum algorithm solving MQb that requires the evaluation of, on average, $O(2^{0.462n})$ quantum gates. To our knowledge this is the fastest algorithm for solving MQb.

6.3.3. MQsoft.

In 2017, NIST shook the cryptographic world by starting a process for standardizing post-quantum cryptography. Sixty-four submissions have been considered for the first round of the on-going NIST Post-Quantum Cryptography (PQC) process. Multivariate cryptography is a classical post-quantum candidate that turns to be the most represented in the signature category. At this stage of the process, it is of primary importance to investigate efficient implementations of the candidates. [17] presents `MQsoft`, an efficient library which permits to implement `HFE`-based multivariate schemes submitted to the NIST PQC process such as *GeMSS*, `Gui` and *DualModeMS*. The library is implemented in `C` targeting Intel 64-bit processors and using `avx2` set instructions. We present performance results for our library and its application to *GeMSS*, `Gui` and *DualModeMS*. In particular, we optimize several crucial parts for these schemes. These include root finding for `HFE` polynomials and evaluation of multivariate quadratic systems in $\mathbb{F}_2$. We propose a new method which accelerates root finding for specific `HFE` polynomials by a factor of two. For *GeMSS* and `Gui`, we obtain a speed-up of a factor between 2 and 19 for the keypair generation, between 1.2 and 2.5 for the signature generation, and between

1.6 and 2 for the verifying process. We have also improved the arithmetic in F_{2^n} by a factor of 4 compared to the NTL library. Moreover, a large part of our implementation is protected against timing attacks.

SECRET Project-Team

7. New Results

7.1. Symmetric cryptology

Participants: Xavier Bonnetain, Christina Boura, Anne Canteaut, Daniel Coggia, Pascale Charpin, Daniel Coggia, Gaëtan Leurent, María Naya Plasencia, Léo Perrin, 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:

- Design of SATURNIN a new lightweight block cipher for authenticated encryption [74], which is resistant to quantum cryptanalysis. SATURNIN has been submitted to the NIST competition for lightweight cryptography, and has been selected for the 2nd round of the competition [0].
- Mixture-differential distinguishers on AES-like ciphers [18].
- Cryptanalysis of the Sbox of the Russian standards, Streebog and Kuznyechik [31], [56]. This work by L. Perrin received the best paper award at *FSE 2019*. Moreover, L. Perrin has been invited to present his results to AFNOR. He is involved in the international standardization processes in symmetric cryptography [50], [86] and has been invited to ISO meetings on this topic.
- The work on the Streebog Sbox has led to a more general study on tools for quantifying anomalies in Sboxes [44].
- Design of BISON, the first concrete block cipher following the whitened swap-or-not construction [46].

7.1.2. MACs and hash functions

The international research effort related to the selection of the new hash function standard SHA-3 has led to many important results and to a better understanding of the security offered by hash functions. However, hash functions are used in a huge number of applications with different security requirements, and also form the building-blocks of some other primitives, like MACs.
Recent results:

- Chosen-prefix collision attack on SHA-1 [52]: A chosen-prefix collision attack is a stronger variant of a collision attack, where an arbitrary pair of challenge prefixes are turned into a collision. Chosen-prefix collisions are usually significantly harder to produce than (identical-prefix) collisions, but the practical impact of such an attack is much larger. G. Leurent and T. Peyrin proposed new techniques to turn collision attacks into chosen-prefix collision attacks, and present such an attack against SHA-1 with complexity between $2^{66.9}$ and $2^{69.4}$ (depending on assumptions about the cost of finding near-collision blocks).
- Design of lighweight MACs from universal hash functions [51]. Many constructions of MACs used in practice (such as GMAC or Poly1305-AES) follow the Wegman-Carter-Shoup construction, which is only secure up to 2^{64} queries with a 128-bit state. S. Duval and G. Leurent proposed new constructions to reach security beyond the birthday bound, and proposed a concrete instantiation, with very good performances on ARM micro-controllers.

[0] https://csrc.nist.gov/CSRC/media/Projects/lightweight-cryptography/documents/round-2/spec-doc-rnd2/saturnin-spec-round2.pdf

7.1.3. 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 [19]. 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 have provided an in-depth analysis of BCT, by studying more closely differentially 4-uniform Sboxes. More recently, C. Boura, L. Perrin and S. Tian have obtained new results on the boomerang uniformity of several constructions of Sboxes [57].
- 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 [21], [58].
- 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 [27]
- Study of the properties of the error-correcting codes associated to differentially 4-uniform Sboxes [26]. Most notably, this work analyzes the relationship between the number of low-weight codewords and the nonlinearity of the corresponding Sbox.
- Study of crooked and weakly-crooked functions [35]: Crooked functions form a family of APN functions whose derivaties take they values in an (affine) hyperplane.
- APN functions with the butterfly construction [22], [34]: the butterfly construction, originally introduced by Perrin et al., is a general construction which includes the only known example of APN permutation operating on an even number of variables. A. Canteaut, L. Perrin and S. Tian have proved that the most recent generalization of this construction does not include any other APN function when the number of variables exceeds six.

7.1.4. Modes of operation and generic attacks

In order to use a block cipher in practice, and to achieve a given security notion, a mode of operation must be used on top of the block cipher. Modes of operation are usually studied through 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:

- Low-memory attacks against the 2-round Even-Mansour construction, using the 3-xor problem [41]: G. Leurent and F. Sibleyras proved that attacking the 2-round Even-Mansour construction with blocksize n is related to the 3-XOR problem with elements on size $2n$. Then, they exhibited the first generic attacks on this construction where both the data and the memory complexity are significantly lower than 2^n.
- Generic attacks against the tweakable FX-construction [55]: F. Sibleyras exhibited a generic attack on the general tweakable iterated FX-construction, which provides an upper-bound on its security. Most notably, for two rounds, this upper bound matches the proof of the particular case of XHX2 by Lee and Lee at Asiacrypt 2018, thus proving for the first time its tightness.
- Modes for authenticated encryption: Besides the design of new lightweight authenticated encryption schemes, we also analyzed some modes of operation in case of release of unverified plaintext (RUP). Indeed, in this setting, an adversary gets separated access to the decryption and verification functionality, and has more power in breaking the scheme. Our results include a forgery attack against the GCM-RUP mode of operation [54], and the design of a new lightweight deterministic scheme, named ANYDAE, which is particularly efficient for short messages, and achieves both conventional security and RUP security [24].
- Generic attacks on hash combiners [15]: G. Leurent and his co-authors analyzed the security of hash combiners, i.e. of procedures that combine two or more hash functions in a way that is hopefully more secure than each of the underlying hash functions, or at least remains secure as long as one of them is secure. They found generic attacks on the XOR combiner, on the concatenation of two Merkle-Damgård hash functions and on the Zipper hash and on the Hash-Twice combiners when they both use Merkle-Damgård hash constructions.

7.2. Code-based cryptography

Participants: Magali Bardet, Kevin Carrier, André Chailloux, Thomas Debris, Matthieu Lequesne, Rocco Mora, Nicolas Sendrier, Jean-Pierre Tillich, Valentin Vasseur.

In recent years, there has been a substantial amount of research on quantum computers. Such computers would be a major threat for all the public-key cryptosystems used in practice, since all these systems rely on the hardness of integer factoring or discrete logarithms, and these problems are easy on a quantum computer. This has prompted NIST to launch a standardization process in 2017 for quantum-safe alternatives to those cryptosystems. This concerns all three major asymmetric primitives, namely public-key encryption schemes, key-exchange protocols and digital signatures. There were 69 valid submissions to this call in November 2017, with numerous lattice-based, code-based and multivariate-cryptography submissions and some submission based either on hashing or on supersingular elliptic curve isogenies. NIST expects to perform multiple rounds of evaluation, over a period of three to five years. The goal of this process is to select a number of acceptable candidate cryptosystems for standardization. The second round of evaluation started in February 2019.

The research of the project-team in this field is focused on the design and cryptanalysis of cryptosystems making use of coding theory. 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). Our recent work on code-based cryptography has to be seen in the context of the recently launched NIST competition for quantum-safe primitives. We have proposed five code-based candidates to the NIST call for the first two primitives, namely public key encryption and key exchange protocols. 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.

We have also been organizing since 2015 a working group held every month or every two months on code-based cryptography that structures the French efforts on this topic: every meeting is attended by most of the groups working in France on this topic (project-team GRACE, University of Bordeaux, University of Limoges, University of Rennes and University of Rouen).

7.2.1. Design of new code-based solutions

The members of the project-team have submitted several candidates to the NIST competition and have designed new code-based primitives.

Recent results:

- Design of a new code-based signature scheme [49]: 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.
- Analysis of the ternary Syndrome Decoding problem [45]: R. Bricout, A. Chailloux, T. Debris and M. Lequesne have performed an algorithmic study of this decoding problem in large weight, which corresponds to the underlying problem in the WAVE signature scheme. Most notably, their study results in an update of the Wave parameters. It also shows that ternary Syndrome Decoding with large weight is a really harder problem than the binary Syndrome Decoding problem, and could have several applications for the design of code-based cryptosystems.

7.2.2. Cryptanalysis of code-based schemes

Recent results:

- Attack against RLCE [48]: 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.
- Analysis of an encryption scheme based on the rank syndrome decoding problem [61]: D. Coggia and A. Couvreur presented an attack against a cryptosystem proposed proposed by Loidreau, which used an intermediary version between Gabidulin codes and LRPC codes. This attack has polynomial time for some parameters of the scheme.
- Decoding algorithm for codes with a non-trivial automorphism group [47]: R. Canto-Torres and JP Tillich presented an algorithm which is able to speed up the decoding of a code with a non-trivial automorphism group. For a certain range of parameters, this results in a decoding that is faster by an exponential factor in the code length when compared to the best algorithms for decoding generic linear codes. This algorithm was then used to break several proposals of public-key cryptosystems based on codes with a non-trivial automorphism group.

7.3. Quantum Information

Participants: Simon Apers, Ivan Bardet, Xavier Bonnetain, Rémi Bricout, André Chailloux, Simona Etinski, Antonio Florez Gutierrez, Shouvik Ghorai, Antoine Grospellier, Lucien Grouès, Anthony Leverrier, Vivien Londe, María Naya Plasencia, Andrea Olivo, Jean-Pierre Tillich, André Schrottenloher, Christophe Vuillot.

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 theses have been defended this year within the project-team on this topic. First, Antoine Grospellier, co-advised by A. Leverrier and O. Fawzi (Ens Lyon), studied efficient decoding algorithms for quantum LDPC codes [13]. 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, co-advised by A. Leverrier and G. Zémor (IMB), worked on the design of better quantum LDPC codes [14]: the main idea is to generalize the celebrated toric code of Kitaev by considering cellulations of manifolds in higher dimensions. A 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.

Lucien Grouès, who did an internship this summer in the project-team, has recently started a PhD with A. Leverrier and O. Fawzi on decoding quantum LDPC codes, and preliminary numerical results have already appeared in [62].

Ivan Bardet joined the project-team as a postdoc in March 2019, and will start a starting research position in 2020. His research focusses on the study of open-system dynamics as well as mixing times of Markovian dissipative evolutions with the goal of better understanding the lifetime of quantum memories.

Recent results:

- Decoding algorithms for Hypergraph Product Codes [62]: this work deals with numerical simulation of several variants of the SMALL-SET-FLIP decoder for hypergraph product codes. While this decoder had already been studied analytically in previous work in the regime of extremely low noise, we are focussing here on understanding its performance for a realistic noise model.
- Towards Low Overhead Magic State Distillation [30]: the major source of overhead in quantum fault-tolerance usually lies in the primitive called magic state distillation which takes a number of noisy versions of a specific quantum state and prepares a new state with less noise. An important question is to understand how efficient this procedure can be. In this work, we prove that magic state distillation can perform much more efficiently than expected when working with quantum systems of large dimension instead of qubits.

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 contributes to this project by studying the security of new key distribution protocols.

Recent results:

- Security proof for two-way continuous-variable quantum key distribution [28]: 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, 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.
- Asymptotic security of continuous-variable quantum key distribution with a discrete modulation [29]: in this work, we establish a lower bound on the secret key rate of a practical quantum key distribution protocol that will be implemented in the context of the H2020 project CiViQ.

7.3.3. *Quantum cryptanalysis of symmetric primitives and quantum algorithms*

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 been awarded an ERC Starting grant for her project named QUASYModo on this topic.

In parallel to this work, S. Apers is developing generic quantum algorithms solving combinatorial problems, notably in graphs. He also recently proposed a unified framework of quantum walk search, that will likely find applications in the context of quantum cryptanalysis.

Recent results:

- Quantum algorithm for the k-XOR problem and for list merging: 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. A. Chailloux, 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 [38], [39].
- Quantum security of AES [17]: In order to determine the post-quantum secuirty margin of AES-256, X. Bonnetain and M. Naya-Plasencia have proposed generalized and quantized versions of the best known cryptanalysis on reduced-round versions of AES-256, including a quantum Demirci-Selçuk meet-in-the-middle attack.
- Quantum attacks without superposition queries : In symmetric cryptanalysis, the model of superposition queries has led to surprising results, but the practical implications of these attacks remain blurry. In contrast, the results obtained so far for a quantum adversary making classical queries only were less impressive. For the first time, M. Naya-Plasencia and A. Schrottenloher, together with A. Hosoyamada and Y. Sasaki, managed to leverage the algebraic structure of some cryptosystems in the context of a quantum attacker limited to classical queries and offline quantum computations. Most notably, they are able to break the Even-Mansour construction in quantum time $\widetilde{\mathcal{O}}(2n/3)$ with $\mathcal{O}(2n/3)$ classical queries and $\mathcal{O}(n^2)$ qubits only.
- Quantum cryptanalysis of CSIDH and Ordinary Isogeny-based Schemes [16]: 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.

- New graph-related quantum algorithms. A first paper presents an approach to improve expansion testing using quantum Fast-Forwarding and growing seed sets [64]. A second paper introduces a graph sparsification algorithm [65], which when combined with existing classical algorithms yields the first quantum speedup for approximating the max cut, min cut, min st-cut, sparsest cut and balanced separator of a graph. Moreover, combining it with a classical Laplacian solver yields a similar speedup for Laplacian solving, for approximating effective resistances, cover times and eigenvalues of the Laplacian, and for spectral clustering.
- Quantum walks: in a first work, S. Apers describes a new quantum algorithm for quantum walk sampling using growing seed sets [42] with applications for st-connectivity and problems related to graph isomorphism. A second work [66] introduces a new quantum walk search framework that unifies and strengthens the existing ones.
- Quantum query lower bounds [59], [60]: Many computational problems, such as finding collisions in a function, are symmetric in their inputs. A. Chailloux showed that for this class of problems, any quantum algorithm can have at most a cubic advantage over the best classical algorithm in the query model, while the previously known bound gave up to 7th root advantage. This result enhances our understanding on the limitations of quantum algorithms.

SPECFUN Project-Team

6. New Results

6.1. 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, whose trailing coefficient (i.e., of order 0) is 1. In [2], 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 published this year.

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

In [8], 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 $\sqrt{p}$.

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

A previous article 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 [3], 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.

6.4. Least common multiple of random integers

In [4], Alin Bostan together with Kilian Raschel (CNRS, Tours) and Alexander Marynych (U. Kyiv, Ukraine) have investigated the least common multiple of random integers. Using a purely probabilistic approach, they derived a criterion for the convergence in distribution as $n\to\infty$ of $f(L_n)/n^{rk}$, for a wide class of multiplicative arithmetic functions f with polynomial growth r, where $L_n(k)$ denotes the least common multiple of k independent random integers with uniform distribution on $\{1,2,...,n\}$. Furthermore, they identified the limit as an infinite product of independent random variables indexed by the prime numbers. Along the way of showing the main results, they computed the (rational) generating function of a trimmed sum of independent geometric laws, which appears in the above infinite product. The latter is directly related to the generating function of a certain max-type diophantine equation, of which they solved a generalized version. The results extend theorems by Erdős and Wintner (1939), Fernández and Fernández (2013) and Hilberdink and Tóth (2016).

6.5. On sequences associated to the invariant theory of rank two simple Lie algebras

In [14], Alin Bostan together with Jordan Tirrell (Washington College, USA) Philadelphia, USA), Bruce W. Westbury (Unversity of Texas at Dallas, USA) and Yi Zhang (Xi'an Jiaotong-Liverpool University, Suzhou, China) studied two families of sequences, listed in the On-Line Encyclopedia of Integer Sequences (OEIS), which are associated to invariant theory of Lie algebras. For the first family, they proved combinatorially that the sequences A059710 and A108307 are related by a binomial transform. Based on this, they presented two independent proofs of a recurrence equation for A059710, which was conjectured by Mihailovs. Besides, they also gave a direct proof of Mihailovs' conjecture by the method of algebraic residues. As a consequence, closed formulae for the generating function of sequence A059710 were obtained in terms of classical Gaussian hypergeometric functions.

6.6. Explicit degree bounds for right factors of linear differential operators

If a linear differential operator with rational function coefficients is reducible, its factors may have coefficients with numerators and denominators of very high degree. When the base field is $\mathbb{C}$, Alin Bostan together with Bruno Salvy (Inria and ENS Lyon) and Tanguy Rivoal (CNRS and U. Grenoble) gave in [13] a completely explicit bound for the degrees of the monic right factors in terms of the degree and the order of the original operator, as well as the largest modulus of the local exponents at all its singularities. As a consequence, if a differential operator L has rational function coefficients over a number field, they obtained degree bounds for its monic right factors in terms of the degree, the order and the height of L, and of the degree of the number field.

6.7. Improved algorithms for left factorial residues

In [11], Alin Bostan together with Vladica Andrejić (University of Belgrade, Serbia) and Milos Tatarevic (CoinList, Alameda, CA) presented improved algorithms for computing the left factorial residues $!p = 0! + 1! + \cdots + (p-1)! \bmod p$. They used these algorithms for the calculation of the residues $!p \bmod p$, for all primes p up to 2^{40}. Their results confirm that Kurepa's left factorial conjecture is still an open problem, as they show that there are no odd primes $p < 2^{40}$ such that p divides $!p$. Additionally, they confirmed that there are no socialist primes p with $5 < p < 2^{40}$.

6.8. A note on gamma triangles and local gamma vectors

Alin Bostan contributed to F. Chapoton's article [5] by writing an appendix, which allowed the author to complete its article. The theme of [5] is the study of simplicial complexes in algebraic combinatorics. A basic invariant is the f-vector that counts faces according to their dimensions. A less understood invariant is the γ-vector, introduced by Gal in 2005. Also in 2005, Chapoton, motivated by the study of the combinatorics of simplicial complexes attached to cluster algebras, considered a refined version of the f-vector. The main aim of [5] is to introduce the analogue in this context of the γ-vector, and a further refinement called the Γ-triangle. The author computed explicitly the Γ-triangle for all the cluster simplicial complexes of irreducible Coxeter groups. Alin Bostan contributed to the proof of an unexpected relation between the Γ-triangles of cluster fans of type $\mathbb{B}$ and $\mathbb{D}$.

6.9. A closed-form formula for the Kullback-Leibler divergence between Cauchy distributions

In the preliminary work [16], Frédéric Chyzak and Frank Nielsen (LIX, Palaiseau and Sony Computer Science Laboratories, Tokyo, Japan) have reported on a closed-form expression for the Kullback-Leibler divergence between Cauchy distributions which involves the calculation of a parametric definite integral with 6 parameters. The formula shows that the Kullback-Leibler divergence between Cauchy densities is always finite and symmetric. This work also serves as a show-case of several methods in computer algebra to the computation of parametrized integrals.

6.10. Big prime field FFT on multi-core processors

In [9], Svyatoslav Covanov, together with Davood Mohajerani, Marc Moreno Maza, and Linxiao Wang (all from ORCCA, Canada), have worked on a multi-threaded implementation of Fast Fourier Transforms over generalized Fermat prime fields. This work extends their previous study realized on graphics processing units to multi-core processors. In this new context, they overcome the less fine control of hardware resources by successively using FFT in support of the multiplication in those fields. They obtain favorable speedup factors (up to 6.9× on a 6-core, 12 threads node, and 4.3× on a 4-core, 8 threads node) of their parallel implementation compared to the serial implementation for the overall application thanks to the low memory footprint and the sharp control of arithmetic instructions of their implementation of generalized Fermat prime fields.

6.11. Martin boundary of killed random walks on isoradial graphs

Alin Bostan contributed to an article by Cédric Boutillier and Kilian Raschel [15], devoted to the study of random walks on isoradial graphs. Contrary to the lattice case, isoradial graphs are not translation invariant, do not admit any group structure and are spatially non-homogeneous. However, Boutillier and Raschel have been able to obtain analogues of a celebrated result by Ney and Spitzer (1966) on the so-called *Martin kernel* (ratio of Green functions started at different points). Alin Bostan provided in the Appendix two different proofs of the fact that some algebraic power series arising in this context have non-negative coefficients.

6.12. Random walks in orthants and lattice path combinatorics

In the second edition of the book [39], original methods were proposed to determine the invariant measure of random walks in the quarter plane with small jumps (size 1), the general solution being obtained via reduction to boundary value problems. Among other things, an important quantity, the so-called *group of the walk*, allows to deduce theoretical features about the nature of the solutions. In particular, when the order of the group is finite and the underlying algebraic curve is of genus 0 or 1, necessary and sufficient conditions have been given for the solution to be rational, algebraic or D-finite (i.e., solution of a linear differential equation). In this framework, a number of difficult open problems related to lattice-path combinatorics are currently being explored by Alin Bostan, Frédéric Chyzak, and Guy Fayolle, both from the theoretical and computer-algebra viewpoints: concrete computation of the criteria, utilization of differential Galois theory, genus greater than 1 (i.e., when some jumps are of size ≥ 2), etc. A recent topic of future research deals with the connections between simple product-form stochastic networks (so-called *Jackson networks*) and explicit solutions of functional equations for counting lattice walks, see [17].

6.13. Quasilinear Average Complexity for Solving Polynomial Systems

How many operations do we need on the average to compute an approximate root of a random Gaussian polynomial system? Beyond Smale's 17th problem that asked whether a polynomial bound is possible, Pierre Lairez has proved in [6] a quasi-optimal bound $(inputsize)^{1+o(1)}$, which improves upon the previously known $(inputsize)^{3/2+o(1)}$ bound. His new algorithm relies on numerical continuation along *rigid continuation paths*. The central idea is to consider rigid motions of the equations rather than line segments in the linear space of all polynomial systems. This leads to a better average condition number and allows for bigger steps. He showed that on the average, one approximate root of a random Gaussian polynomial system of n equations of degree at most D in $n+1$ homogeneous variables can be computed with $O(n^5 D^2)$ continuation steps. This is a decisive improvement over previous bounds, which prove no better than $\sqrt{2}^{\min(n,D)}$ continuation steps on the average.

In 2019, the article has been accepted in the Journal of the AMS.

6.14. Computing the Volume of Compact Semi-Algebraic Sets

In [10], Pierre Lairez, Mohab Safey El Din and Marc Mezzarobba join a unique set of expertise in symbolic integration, real algebraic geometry and numerical integration to tackle a problem as old as Babylonian mathematics: the computation of volumes.

Let $S \subset R^n$ be a compact basic semi-algebraic set defined as the real solution set of multivariate polynomial inequalities with rational coefficients. They design an algorithm which takes as input a polynomial system defining S and an integer $p \geq 0$ and returns the n-dimensional volume of S at absolute precision 2^{-p}.

Their algorithm relies on the relationship between volumes of semi-algebraic sets and periods of rational integrals. It makes use of algorithms computing the Picard-Fuchs differential equation of appropriate periods, properties of critical points, and high-precision numerical integration of differential equations.

The algorithm runs in essentially linear time with respect to p. This improves upon the previous exponential bounds obtained by Monte-Carlo or moment-based methods.

6.15. Densities of Stieltjes moment sequences for pattern-avoiding permutations

A small subset of combinatorial sequences have coefficients that can be represented as moments of a nonnegative measure on $[0, \infty)$. Such sequences are known as *Stieltjes moment sequences*. They have a number of useful properties, such as log-convexity, which in turn enables one to rigorously bound their growth constant from below.

In [12], Alin Bostan together with Andrew Elvey Price, Anthony Guttmann and Jean-Marie Maillard, studied some classical sequences in enumerative combinatorics, denoted $Av(\mathcal{P})$, and counting permutations of $\{1, 2, ..., n\}$ that avoid some given pattern $\mathcal{P}$. For increasing patterns $\mathcal{P} = (12...k)$, they showed that the corresponding sequences, $Av(123...k)$, are Stieltjes moment sequences, and explicitly determined the underlying density function, either exactly or numerically, by using the Stieltjes inversion formula as a fundamental tool.

They showed that the densities for $Av(1234)$ and $Av(12345)$, correspond to an order-one linear differential operator acting on a classical modular form given as a pullback of a Gaussian ${}_2F_1$ hypergeometric function, respectively to an order-two linear differential operator acting on the square of a classical modular form given as a pullback of a ${}_2F_1$ hypergeometric function. Moreover, these density functions are closely, but non-trivially, related to the density attached to the distance traveled by a walk in the plane with $k - 1$ unit steps in random directions.

As a bonus, they studied the challenging case of the $Av(1324)$ sequence and gave compelling numerical evidence that this too is a Stieltjes moment sequence. Accepting this, they proved new lower bounds on the growth constant of this sequence, which are stronger than existing bounds. A further unproven assumption leads to even better bounds, which can be extrapolated to give a good estimate of the (unknown) growth constant.

CAIRN Project-Team

6. New Results

6.1. Reconfigurable Architecture and Hardware Accelerator Design

6.1.1. Algorithmic Fault Tolerance for Timing Speculative Hardware

Participants: Thibaut Marty, Tomofumi Yuki, Steven Derrien.

We have been working on timing speculation, also known as overclocking, to increase the computational throughput of accelerators. However, aggressive overclocking introduces timing errors, which may corrupt the outputs to unacceptable levels. It is extremely challenging to ensure that no timing errors occur, since the probability of such errors happening depends on many factors including the temperature and process variation. Thus, aggressive timing speculation must be coupled with a mechanism to verify that the outputs are correctly computed. Our previous result demonstrated that the use of inexpensive checks based on algebraic properties of the computation can drastically reduce the cost of verifying that overclocking did not produce incorrect outputs. This has allowed the accelerator to significantly boost its throughput with little area overhead.

One weakness coming from the use of algebraic properties is that the inexpensive check is not strictly compatible with floating-point arithmetic that is not associative. This was not an issue with our previous work that targeted convolutional neural networks, which typically use fixed-point (integer) arithmetic. Our on-going work aims to extend our approach to floating-point arithmetic by using extended precision to store intermediate results, known as Kulisch accumulators. At first glance, use of extended precision that covers the full exponent range of floating-point may look costly. However, the design space of FPGAs is complex with many different trade-offs, making the optimal design highly context dependent. Our preliminary results indicate that the use of extended precision may not be any more costly than implementing the computation in floating point.

6.1.2. Adaptive Dynamic Compilation for Low-Power Embedded Systems

Participants: Steven Derrien, Simon Rokicki.

Previous works on Hybrid-DBT have demonstrated that using Dynamic Binary Translation, combined with low-power in-order architecture, enables an energy-efficient execution of compute-intensive kernels. In [33], we address one of the main performance limitations of Hybrid-DBT: the lack of speculative execution. We study how it is possible to use memory dependency speculation during the DBT process. Our approach enables fine-grained speculation optimizations thanks to a combination of hardware and software mechanisms. Our results show that our approach leads to a geo-mean speed-up of 10% at the price of a 7% area overhead. In [49], we summarize the current state of the Hybrid-DBT project and display our last results about the performance and the energy efficiency of the system. The experimental results presented here show that, for compute-intensive benchmarks, Hybrid-DBT can deliver the same performance level than a 3-issue OoO core, while consuming three times less energy. Finally, in [34], we investigate security issues caused by the use of speculation in DBT-based systems. We demonstrate that, even if those systems use in-order micro-architectures, the DBT layer optimizes binaries and speculates on the outcome of some branches, leading to security issues similar to the Spectre vulnerability. We demonstrate that both the NVidia Denver architecture and the Hybrid-DBT platform are subject to such vulnerability. However, we also demonstrate that those systems can easily be patched, as the DBT is done in software and has fine-grained control over the optimization process.

6.1.3. What You Simulate Is What You Synthesize: Designing a Processor Core from C++ Specifications

Participants: Simon Rokicki, Davide Pala, Joseph Paturel, Olivier Sentieys.

Designing the hardware of a processor core as well as its verification flow from a single high-level specification would provide great advantages in terms of productivity and maintainability. In [32] (a preliminary version also in [42]), we highlight the gain of starting from a unique high-level synthesis and simulation C++ model to design a processor core implementing the RISC-V Instruction Set Architecture (ISA). The specification code is used to generate both the hardware target design through High-Level Synthesis as well as a fast and cycle-accurate bit-accurate simulator of the latter through software compilation. The object oriented nature of C++ greatly improves the readability and flexibility of the design description compared to classical HDL-based implementations. Therefore, the processor model can easily be modified, expanded and verified using standard software development methodologies. The main challenge is to deal with C++ based synthesizable specifications of core and uncore components, cache memory hierarchy, and synchronization. In particular, the research question is how to specify such parallel computing pipelines with high-level synthesis technology and to demonstrate that there is a potential high gain in design time without jeopardizing performance and cost. Our experiments demonstrate that the core frequency and area of the generated hardware are comparable to existing RTL implementations.

6.1.4. Accelerating Itemset Sampling on FPGA

Participants: Mael Gueguen, Olivier Sentieys.

Finding recurrent patterns within a data stream is important for fields as diverse as cybersecurity or e-commerce. This requires to use pattern mining techniques. However, pattern mining suffers from two issues. The first one, known as ”pattern explosion”, comes from the large combinatorial space explored and is the result of too many patterns outputted to be analyzed. Recent techniques called output space sampling solve this problem by outputting only a sampled set of all the results, with a target size provided by the user. The second issue is that most algorithms are designed to operate on static datasets or low throughput streams. In [24], we propose a contribution to tackle both issues, by designing an FPGA accelerator for pattern mining with output space sampling. We show that our accelerator can outperform a state-of-the-art implementation on a server class CPU using a modest FPGA product. This work is done in collaboration with A. Termier from the Lacodam team at Inria.

6.1.5. Hardware Accelerated Simulation of Heterogeneous Platforms

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

When considering designing heterogeneous multicore 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 multicore architectures. These limitations motivate the use of hardware to accelerate the simulation of heterogeneous multicore, and in particular of FPGA components. We study an approach for designing such systems based on performance models through combining accelerator and processor core models. These models are implemented in the HAsim/LEAP infrastructure. In [22], we propose a methodology for building performance models of accelerators and describe the defined design flow.

6.1.6. Fault-Tolerant Scheduling onto Multicore embedded Systems

Participants: Emmanuel Casseau, Minyu Cui, Petr Dobias, Lei Mo, Angeliki Kritikakou.

Demand on multiprocessor 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, multiprocessor 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 map and schedule tasks onto homogeneous faulty processors.

We consider two approaches. The first approach deals with task mapping onto processors at compile time. Our goal is to guarantee both reliability and hard real-time constraints with low-energy consumption. Task duplication is assessed and duplication is performed if expected reliability of a task is not met. This work concurrently decides duplication of tasks, the task execution frequency and task allocation to minimize the energy consumption of a multicore platform with Dynamic Voltage and Frequency Scaling (DVFS) capabilities. The problem is initially formulated as Integer Non-Linear Programming and equivalently transformed to a Mixed Integer Linear Programming problem to be optimally solved. The proposed approach provides a good trade-off between energy consumption and reliability. The second approach deals with mapping and scheduling tasks at runtime. The application context is CubeSats. CubeSats operate in harsh space environment and they are exposed to charged particles and radiations, which cause transient faults. To make CubeSats fault tolerant, we propose to take advantage of their multicore architecture. We propose two online algorithms, which schedule all tasks on board of a CubeSat, detect faults and take appropriate measures (based on task replication) in order to deliver correct results. The first algorithm considers all tasks as aperiodic tasks and the second one treats them as aperiodic or periodic tasks. Their performances vary, particularly when the number of processors is low, and a choice is subject to a trade-off between the rejection rate and the energy consumption. This work is done in collaboration with Oliver Sinnen, PARC Lab., the University of Auckland.

6.1.7. Run-Time Management on Multicore Platforms

Participant: Angeliki Kritikakou.

In time-critical systems, run-time adaptation is required to improve the performance of time-triggered execution, derived based on Worst-Case Execution Time (WCET) of tasks. By improving performance, the systems can provide higher Quality-of-Service, in safety-critical systems, or execute other best-effort applications, in mixed-critical systems. To achieve this goal, we propose a parallel interference-sensitive run-time adaptation mechanism that enables a fine-grained synchronisation among cores [37]. Since the run-time adaptation of offline solutions can potentially violate the timing guarantees, we present the Response-Time Analysis (RTA) of the proposed mechanism showing that the system execution is free of timing-anomalies. The RTA takes into account the timing behavior of the proposed mechanism and its associated WCET. To support our contribution, we evaluate the behavior and the scalability of the proposed approach for different application types and execution configurations on the 8-core Texas Instruments TMS320C6678 platform. The obtained results show significant performance improvement compared to state-of-the-art centralized approaches.

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

Participants: Olivier Sentieys, Angeliki Kritikakou, Lei Mo.

Asymmetric Multicore Processors (AMP) are a very promising architecture to deal efficiently with the wide diversity of applications. In real-time application domains, in-time approximated results are preferred to accurate – but too late – results. In [28], we propose a deployment approach that exploits the heterogeneity provided by AMP architectures and the approximation tolerance provided by the applications, so as to increase as much as possible the quality of the results under given energy and timing constraints. Initially, an optimal approach is proposed based on the problem linearization and decomposition. Then, a heuristic approach is developed based on iteration relaxation of the optimal version. The obtained results show 16.3% reduction in the computation time for the optimal approach compared to conventional optimal approaches. The proposed heuristic approach is about 100 times faster at the cost of a 29.8% QoS degradation in comparison with the optimal solution.

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

Participants: Angeliki Kritikakou, Lei Mo.

Cyber-Physical Systems (CPS), as a particular case of distributed systems, raise new challenges, because of the heterogeneity and other properties traditionally associated with Wireless Sensor and Actuator Networks (WSAN), including shared sensing, acting and real-time computing. In CPS, mobile actuators can enhance system's flexibility and scalability, but at the same time incur complex couplings in the scheduling and controlling of the actuators. In [19], we propose a novel event-driven method aiming at satisfying a required level of control accuracy and saving energy consumption of the actuators, while guaranteeing a bounded action delay. We formulate a joint-design problem of both actuator scheduling and output control. To solve this problem, we propose a two-step optimization method. In the first step, the problem of actuator scheduling and action time allocation is decomposed into two subproblems. They are solved iteratively by utilizing the solution of one in the other. The convergence of this iterative algorithm is proved. In the second step, an on-line method is proposed to estimate the error and adjust the outputs of the actuators accordingly. Through simulations and experiments,we demonstrate the effectiveness of the proposed method. In addition, many of the real-time tasks of CPS can be executed in an imprecise way. Such systems accept an approximate result as long as the baseline Quality-of-Service (QoS) is satisfied and they can execute more computations to yield better results, if more system resources are available. These systems are typically considered under the Imprecise Computation (IC) model, achieving a better tradeoff between QoS and limited system resources. However, determining a QoS-aware mapping of these real-time IC-tasks onto the nodes of a CPS creates a set of interesting problems. In [18], we firstly propose a mathematical model to capture the dependency, energy and real-time constraints of IC-tasks, as well as the sensing, acting, and routing in the CPS. The problem is formulated as a Mixed-Integer Non-Linear Programming (MINLP) due to the complex nature of the problem. Secondly, to efficiently solve this problem, we provide a linearization method that results in a Mixed-Integer Linear Programming (MILP) formulation of our original problem. Finally, we decompose the transformed problem into a task allocation subproblem and a task adjustment subproblem, and, then, we find the optimal solution based on subproblem iteration. Through the simulations, we demonstrate the effectiveness of the proposed method. Last, but not least, wireless charging can provide dynamic power supply for CPS. Such systems are typically considered under the scenario of Wireless Rechargeable Sensor Networks (WRSNs). With the use of Mobile Chargers (MCs), the flexibility of WRSNs is further enhanced. However, the use of MCs poses several challenges during the system design. The coordination process has to simultaneously optimize the scheduling, the moving time and the charging time of multiple MCs, under limited system resources (e.g., time and energy). Efficient methods that jointly solve these challenges are generally lacking in the literature. In [17], we address the multiple MCs coordination problem under multiple system requirements. Firstly, we aim at minimizing the energy consumption of MCs, guaranteeing that every sensor will not run out of energy. We formulate the multiple MCs coordination problem as a mixed-integer linear programming and derive a set of desired network properties. Secondly, we propose a novel decomposition method to optimally solve the problem, as well as to reduce the computation time. Our approach divides the problem into a subproblem for the MC scheduling and a subproblem for the MC moving time and charging time, and solves them iteratively by utilizing the solution of one into the other. The convergence of the proposed method is analyzed theoretically. Simulation results demonstrate the effectiveness and scalability of the proposed method in terms of solution quality and computation time.

6.1.10. Fault-Tolerant Microarchitectures

Participants: Joseph Paturel, Angeliki Kritikakou, Olivier Sentieys.

As transistors scale down, processors are more vulnerable to radiation that can cause multiple transient faults in function units. Rather than excluding these units from execution, performance overhead of VLIW processors can be reduced when fault-free components of these affected units are still used. In [30], the function units are enhanced with coarse-grained fault detectors. A re-scheduling of the instructions is performed at run-time to use not only the healthy function units, but also the fault-free components of the faulty function units. The scheduling window of the proposed mechanism covers two instruction bundles, which makes it suitable to explore mitigation solutions in the current and in the next instruction execution. Experiments show that the proposed approach can mitigate a large number of faults with low performance and area overheads. In addition, technology scaling can cause transient faults with long duration. In this case, the affected function unit is usually considered as faulty and is not further used. To reduce this performance degradation, we proposed a hardware mechanism to (i) detect the faults that are still active during execution and (ii) re-schedule the instructions to use the fault-free components of the affected function units [31]. When the fault faints, the affected function unit components can be reused. The scheduling window of the proposed mechanism is two instruction bundles being able to exploit function units of both the current and the next instruction execution.

The results show multiple long-duration fault mitigation can be achieved with low performance, area, and power overhead.

Simulation-based fault injection is commonly used to estimate system vulnerability. Existing approaches either partially model the studied system's fault masking capabilities, losing accuracy, or require prohibitive estimation times. Our work proposes a vulnerability analysis approach that combines gate-level fault injection with microarchitecture-level Cycle-Accurate and Bit-Accurate simulation, achieving low estimation time. Faults both in sequential and combinational logic are considered and fault masking is modeled at gate-level, microarchitecture-level and application-level, maintaining accuracy. Our case-study is a RISC-V processor. Obtained results show a more than 8% reduction in masked errors, increasing more than 55% system failures compared to standard fault injection approaches. This work is currently under review.

6.1.11. Fault-Tolerant Networks-on-Chip

Participants: Romain Mercier, Cédric Killian, Angeliki Kritikakou, Daniel Chillet.

Network-on-Chip has become the main interconnect in the multicore/manycore era since the beginning of this decade. However, these systems become more sensitive to faults due to transistor shrinking size. In parallel, approximate computing appears as a new computation model for applications since several years. The main characteristic of these applications is to support the approximation of data, both for computations and for communications. To exploit this specific application property, we develop a fault-tolerant NoC to reduce the impact of faults on the data communications. To address this problem, we consider multiple permanent faults on router which cannot be managed by Error-Correcting Codes (ECCs) and we propose a bit-shuffling method to reduce the impact of faults on Most Significant Bits (MSBs), hence permanent faults only impact Low Significant Bits (LSBs) instead of MSBs reducing the errors impact. We evaluated the proposed method for data mining benchmark and we show that our proposal can lead to 73.04% reduction on the clustering error rate and 84.64% reduction on the mean centroid Mean Square Error (MSE) for 3-bit permanent faults which affect MSBs on 32-bit words with a limited area cost. This work is currently under review for an international conference.

6.1.12. Improving the Reliability of Wireless Network-on-Chip (WiNoC)

Participants: Joel Ortiz Sosa, Olivier Sentieys, Cédric Killian.

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, standard WiNoC approaches are vulnerable to multi-path interference introduced by on-chip physical structures. To overcome such parasitic phenomenon, we first proposed a Time-Diversity Scheme (TDS) to enhance the reliability of on-chip wireless links using a realistic wireless channel model. We then proposed an adaptive digital transceiver, which enhances communication reliability under different wireless channel configurations in [39]. Based on the same realistic channel model, we investigated the impact of using some channel correction techniques. Experimental results show that our approach significantly improves Bit Error Rate (BER) under different wireless channel configurations. Moreover, our transceiver is designed to be adaptive, which allows for wireless communication links to be established in conditions where this would not be possible for standard transceiver architectures. The proposed architecture, designed using a 28-nm FDSOI technology, consumes only 3.27 mW for a data rate of 10 Gbit/s and has a very small area footprint. We also proposed a low-power, high-speed, multi-carrier reconfigurable transceiver based on Frequency Division Multiplexing (FDM) to ensure data transfer in future Wireless NoCs in [38]. The proposed transceiver supports a medium access control method to sustain unicast, broadcast and multicast communication patterns, providing dynamic data exchange among wireless nodes. Designed using a 28-nm FDSOI technology, the transceiver only consumes 2.37 mW and 4.82 mW in unicast/broadcast and multicast modes, respectively, with an area footprint of 0.0138 mm^2.

6.1.13. Error Mitigation in Nanophotonic Interconnect

Participants: Jaechul Lee, Cédric Killian, Daniel Chillet.

The energy consumption of manycore is dominated by data movements, which calls for energy-efficient and high-bandwidth interconnects. Integrated optics is promising technology to overcome the bandwidth limitations of electrical interconnects. However, it suffers from high power overhead related to low efficiency lasers, which calls for the use of approximate communications for error tolerant applications. In this context, in [26] we investigate the design of an Optical NoC supporting the transmission of approximate data. For this purpose, the least significant bits of floating point numbers are transmitted with low power optical signals. A transmission model allows estimating the laser power according to the targeted BER and a micro-architecture allows configuring, at run-time, the number of approximated bits and the laser output powers. Simulation results show that, compared to an interconnect involving only robust communications, approximations in the optical transmissions lead to a laser power reduction up to 42% for image processing application with a limited degradation at the application level.

6.2. Compilation and Synthesis for Reconfigurable Platform

6.2.1. Compile Time Simplification of Sparse Matrix Code Dependences

Participant: Tomofumi Yuki.

In [29], we developed a combined compile-time and runtime loop-carried dependence analysis of sparse matrix codes and evaluated its performance in the context of wavefront parallellism. Sparse computations incorporate indirect memory accesses such as x[col[j]] whose memory locations cannot be determined until runtime. The key contributions are two compile-time techniques for significantly reducing the overhead of runtime dependence testing: (1) identifying new equality constraints that result in more efficient runtime inspectors, and (2) identifying subset relations between dependence constraints such that one dependence test subsumes another one that is therefore eliminated. New equality constraints discovery is enabled by taking advantage of domain-specific knowledge about index arrays, such as col[j]. These simplifications lead to automatically-generated inspectors that make it practical to parallelize such computations. We analyze our simplification methods for a collection of seven sparse computations. The evaluation shows our methods reduce the complexity of the runtime inspectors significantly. Experimental results for a collection of five large matrices show parallel speedups ranging from 2x to more than 8x running on a 8-core CPU.

6.2.2. Study of Polynomial Scheduling

Participant: Tomofumi Yuki.

We have studied the Handelman's theorem used for polynomial scheduling, which resembles the Farkas' lemma for affine scheduling. Theorems from real algebraic geometry and polynomial optimization show that some polynomials have Handelman representations when they are non-negative on a domain, instead of strictly positive as stated in Handelman's theorem. The global minimizers of a polynomial must be at the boundaries of the domain to have such a representation with finite bounds on the degree of monomials. This creates discrepancies in terms of polynomials included in the exploration space with a fixed bound on the monomial degree. Our findings give an explanation to our failed attempt to apply polynomial scheduling to Index-Set Splitting: we were precisely trying to find polynomials with global minimizers at the interior of a domain.

6.2.3. Optimizing and Parallelizing compilers for Time-Critical Systems

Participant: Steven Derrien.

6.2.3.1. Contentions-Aware Task-Level Parallelization

Accurate WCET analysis for multicores is challenging due to concurrent accesses to shared resources, such as communication through bus or Network on Chip (NoC). Current WCET techniques either produce pessimistic WCET estimates or preclude conflicts by constraining the execution, at the price of a significant hardware under-utilization. Most existing techniques are also restricted to independent tasks, whereas real-time workloads will probably evolve toward parallel programs. The WCET behavior of such parallel programs is even more challenging to analyze because they consist of *dependent* tasks interacting through complex synchronization/communication mechanisms. In [36], we propose a scheduling technique that jointly selects

Scratchpad Memory (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 technique on streaming applications and synthetic task-graphs. The overlapping of communications with computations allows the length of generated schedules to be reduced by 4% on average on streaming applications, with a maximum of 16%, and by 8% on average for synthetic task graphs. We further show on a case study that generated schedules can be implemented with low overhead on a predictable multicore architecture (Kalray MPPA).

6.2.3.2. WCET-Aware Parallelization of Model-Based Applications for Multicores

Parallel architectures are nowadays increasingly used in embedded time-critical systems. The 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 multicores. 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 [20] we present the overall results of the project, a compiler tool-flow for automated parallelization of model-based real-time software, which addresses the shortcomings of multi-core architectures in real-time systems. The flow is demonstrated using a model-based Terrain Awareness and Warning Systems (TAWS) and an edge detection algorithm from the image-processing domain. Model-based applications are first transformed into real-time C code and from there into a well-predictable parallel C program. Tight bounds for the Worst-Case Execution Time (WCET) of the parallelized program can be determined using an integrated multicore WCET analysis. Thanks to the use of an architecture description language, the general approach is applicable to a wider range of target platforms. An experimental evaluation for a research architecture with network-on-chip (NoC) interconnect shows that the parallel WCET of the TAWS application can be improved by factor 1.77 using the presented compiler tools.

6.2.3.3. WCET oriented Iterative compilation

Static Worst-Case Execution Time (WCET) estimation techniques operate upon the binary code of a program in order to provide the necessary input for schedulability analysis techniques. Compilers used to generate this binary code include tens of optimizations, that can radically change the flow information of the program. Such information is hard to maintain across optimization passes and may render automatic extraction of important flow information, such as loop bounds, impossible. Thus, compiler optimizations, especially the sophisticated optimizations of mainstream compilers, are typically avoided. In this work, published in [23], we explore for the first time iterative-compilation techniques that reconcile compiler optimizations and static WCET estimation. We propose a novel learning technique that selects sequences of optimizations that minimize the WCET estimate of a given program. We experimentally evaluate the proposed technique using an industrial WCET estimation tool (AbsInt aiT) over a set of 46 benchmarks from four different benchmarks suites, including reference WCET benchmark applications, image processing kernels and telecommunication applications. Experimental results show that WCET estimates are reduced on average by 20.3% using the proposed technique,as compared to the best compiler optimization level applicable.

6.2.4. Towards Generic and Scalable Word-Length Optimization

Participants: Van-Phu Ha, Tomofumi Yuki, Olivier Sentieys.

Fixed-Point arithmetic is widely used for implementing Digital Signal Processing (DSP) systems on electronic devices. Since initial specifications are often written using floating-point arithmetic, conversion to fixed-point is a recurring step in hardware design. The primary objective of this conversion is to minimize the cost (energy and/or area) while maintaining an acceptable level of quality at the output. In Word-Length Optimization (WLO), each variable/operator may be assigned a different fixed-point encoding, which means that the design space grows exponentially as the number of variables increases. This is especially true when targeting hardware accelerators implemented in FPGA or ASIC. Thus, most approaches for WLO involve heuristic search algorithms. In [25] (a preliminary version also in [41]), we propose a method to improve the scalability of Word-Length Optimization (WLO) for large applications that use complex quality metrics such as Structural

Similarity (SSIM). The input application is decomposed into smaller kernels to avoid uncontrolled explosion of the exploration time, which is known as noise budgeting. The main challenge addressed in this paper is how to allocate noise budgets to each kernel. This requires capturing the interactions across kernels. The main idea is to characterize the impact of approximating each kernel on accuracy/cost through simulation and regression. Our approach improves the scalability while finding better solutions for Image Signal Processor pipeline.

In [27], we propose an analytical approach to study the impact of floating-point (FlP) precision variation on the square root operation, in terms of computational accuracy and performance gain. We estimate the round-off error resulting from reduced precision. We also inspect the Newton Raphson algorithm used to approximate the square root in order to bound the error caused by algorithmic deviation. Consequently, the implementation of the square root can be optimized by fittingly adjusting its number of iterations with respect to any given FlP precision specification, without the need for long simulation times. We evaluate our error analysis of the square root operation as part of approximating a classic data clustering algorithm known as K-means, for the purpose of reducing its energy footprint. We compare the resulting inexact K-means to its exact counterpart, in the context of color quantization, in terms of energy gain and quality of the output. The experimental results show that energy savings could be achieved without penalizing the quality of the output (e.g., up to 41.87% of energy gain for an output quality, measured using structural similarity, within a range of [0.95,1]).

6.2.5. Optimized Implementations of Constant Multipliers for FPGAs

Participant: Silviu-Ioan Filip.

The multiplication by a constant is a frequently used arithmetic operation. To implement it on Field Programmable Gate Arrays(FPGAs), the state of the art offers two completely different methods: one relying on bit shifts and additions/subtractions, and another one using look-up tables and additions. So far, it was unclear which method performs best for a given constant and input/output data types. The main contribution of the work published in [40] is a thorough comparison of both methods in the main application contexts of constant multiplication: filters, signal-processing transforms, and elementary functions. Most of the previous state of the art addresses multiplication by an integer constant. This work shows that, in most of these application contexts, a formulation of the problem as the multiplication by a real constant allows for more efficient architectures. Another contribution is a novel extension of the shift-and-add method to real constants. For that, an integer linear programming (ILP) formulation is proposed, which truncates each component in the shift-and-add network to a minimum necessary word size that is aligned with the approximation error of the coefficient. All methods are implemented within the open-source FloPoCo framework.

6.2.6. Optimal Multiplierless FIR Filter Design

Participant: Silviu-Ioan Filip.

The hardware optimization of direct form finite impulse response (FIR) filters has been a topic of research for the better part of the last four decades and is still garnering significant research and industry interest. In [48], we present two novel optimization methods based on integer linear programming (ILP) that minimize the number of adders used to implement a direct/transposed FIR filter adhering to a given frequency specification. The proposed algorithms work by either fixing the number of adders used to implement the products (multiplier block adders) or by bounding the adder depth (AD) used for these products. The latter can be used to design filters with minimal AD for low power applications. In contrast to previous multiplierless FIR approaches, the methods introduced here ensure adder count optimality. To demonstrate their effectiveness, we perform several experiments using established design problems from the literature, showing superior results.

6.2.7. Application-specific arithmetic in high-level synthesis tools

Participant: Steven Derrien.

In [50], we have shown that the use of non-conventional implementation for floating-point arithmetic can bring significant benefits when used in the context of High-Level Synthesis. We are currently building on these preliminary results to show that it is possible to implement accelerators using exact floating-point arithmetic for similar performance/area cost than standard floating-point operators implementations. Our approach builds on Kulish's approach to implement floating-point adders, and targets dense Matrix Products kernels (GEM3 like) accelerators on FPGAs.

6.3. Applications

6.3.1. SmartSense

Participants: Nicolas Roux, Olivier Sentieys.

Developing smarter and greener buildings has been an expanding field of research over the last decades. One of the essential requirements for energy utilities is the knowledge of power consumption patterns at the single-appliance level. To estimate these patterns without using an individual power meter for each appliance, Non-Intrusive Load Monitoring (NILM) consists in disaggregating electrical loads by examining the appliance specific power consumption signature within the aggregated load single measurement. Therefore, the method is considered non-intrusive since the data are collected from a single electrical panel outside of the monitored building. Thus, NILM has been a very active field of research with renewed interest over the last years.

Therefore, knowing the plug-level power consumption of each appliance in a building can lead to drastic savings in energy consumption. In [35], we have addressed the issue of NILM inaccuracy in the context of industrial or commercial buildings, by combining data from a low-cost, general-purpose, wireless sensor network. We have proposed a novel approach based on a simplex solver to estimate the power load values of the steady states on sliding windows of data with varying size. We have shown the principle of the approach and demonstrated its interest, limited complexity, and ease of use.

CAMUS Project-Team

7. New Results

7.1. The Polyhedral Model Beyond Loops

Participants: Salwa Kobeissi, Philippe Clauss.

There may be a huge gap between the statements outlined by programmers in a program source code and instructions that are actually performed by a given processor architecture when running the executable code. This gap is due to the way the input code has been interpreted, translated and transformed by the compiler and the final processor hardware. Thus, there is an opportunity for efficient optimization strategies, that are dedicated to specific control structures and memory access patterns, to be applied as soon as the actual runtime behavior has been discovered, even if they could not have been applied on the original source code.

We develop this idea by identifying code excerpts that behave as polyhedral-compliant loops at runtime, while not having been outlined at all as loops in the original source code. In particular, we are interested in recursive functions whose runtime behavior can be modeled as polyhedral loops. Therefore, the scope of this study exclusively includes 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 (NLR) algorithm [3], performing the affine loop modeling of the original program runtime behavior, which is then used to generate an equivalent iterative program, finally optimized using the polyhedral compiler Polly. We present some preliminary results showing that 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 published at the 9th International Workshop on Polyhedral Compilation Techniques [22].

7.2. New release of Apollo

Participants: Muthena Abdul-Wahab, Philippe Clauss.

Apollo has been updated to use LLVM/Clang version 6.0.1. The unmodified sources are now included, as tar-files, in the APOLLO distribution.

Regarding the build system:

- All components of APOLLO are now installed into the installation directory. Once installed, APOLLO does not need the build directory to be kept.
- The RPATH on APOLLO libraries has been set to the installation directory. This allows APOLLO to be run without having to set up library paths.
- APOLLO_BUILD_JOBS has been introduced to specify the maximum number of build jobs to use. The replaces NB_JOBS which is still supported but deprecated.
- The sources for external dependencies are now included in the APOLLO distribution. They are no longer downloaded during a build.
- A new build target 'check' has been added to run the testsuite. This is supported by Makefiles ('make check') and Ninja ('ninja check').
- The build type (Debug/Release) for LLVM/Clang is now the same as the rest of APOLLO. New build variable APOLLO_LLVM_BUILD_TYPE can be used to specify a separate build type for LLVM/Clang.

Regarding bug fixes:

- Valid code using floating point types (float or double) could make APOLLO stop with an message about unsupported scalars. This has been fixed by removing the Loop Invariant Code Motion (LICM) pass in such cases, preventing floating-point scalars to be generated.
- Code containing try-catch blocks could make APOLLO crash. This has been fixed.
- Dynamic loop bounds were no more instrumented and interpolated. This has been fixed.

7.3. Uniform Random Sampling in Polyhedra

Participant: Philippe Clauss.

We propose a method for generating uniform samples among a domain of integer points defined by a polyhedron in a multi-dimensional space. The method extends to domains defined by parametric polyhedra, in which a subset of the variables are symbolic. We motivate this work by a list of applications for the method in computer science. The proposed method relies on polyhedral ranking functions, as well as a recent inversion method for them, named *trahrhe* expressions. This work has been accomplished in collaboration with Benoît Meister from Reservoir Labs, New York, USA, and has been published at the 10th International Workshop on Polyhedral Compilation Techniques, January 2020.

7.4. Runtime Multi-Versioning and Specialization

Participant: Philippe Clauss.

We have developped an extension of APOLLO that implements code multi-versioning and specialization to optimize and parallelize loop kernels that are invoked many times with varying parameters. These parameters may influence the code structure, the touched memory locations, the workload, and the runtime performance. They may also impact the validity of the parallelizing and optimizing polyhedral transformations that are applied on-the-fly.

For a target loop kernel and its associated parameters, a different optimizing and parallelizing transformation is evaluated at each invocation, among a finite set of transformations (multi-versioning and specialization). The best performing transformed code version is stored and indexed using its associated parameters. When every optimizing transformation has been evaluated, the best performing code version regarding the current parameters, which has been stored, is relaunched at next invocations (memoization).

This work has been accomplished in collaboration with Raquel Lazcano and Eduardo Juarez of the Universidad Politécnica de Madrid, Spain, and has been published at the ACM SIGPLAN 2020 International Conference on Compiler Construction (CC 2020).

7.5. AutoParallel: Automatic parallelization and distributed execution of affine loop nests in Python

Participant: Philippe Clauss.

The last improvements in programming languages and models have focused on simplicity and abstraction; leading Python to the top of the list of the programming languages. However, there is still room for improvement when preventing users from dealing directly with distributed and parallel computing issues. We propose AutoParallel, a Python module to automatically find an appropriate task-based parallelisation 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 tasks' 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 intermediate-level programming skills can scale up an application to hundreds of cores.

This work has been accomplished in collaboration with Cristian Ramon-Cortes, Ramon Amela, Jorge Ejarque and Rosa M. Badia of the Barcelona Supercomputing Center (BSC), Spain. A journal paper is in preparation.

7.6. Combining Locking and Data Management Interfaces

Participants: Jens Gustedt, Daniel Salas.

Handling data consistency in parallel and distributed settings is a challenging task, in particular if we want to allow for an easy to handle asynchronism between tasks. Our publication [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 6.3 . In previous work it has demonstrated its efficiency for a large variety of platforms.

In the framework of the ASNAP project 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, for each individual data object we conserve the same modification order as the sequential algorithm. 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 [16] for first results.

In the context of the Prim'Eau project (see 9.1.2) 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. A first step to this goal has been a specific decomposition of geological data, see [21].

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. This year, Daniel has successfully defended his thesis, see [7].

7.7. Granularity Control for Parallel Programs

Participant: Arthur Charguéraud.

Arthur Charguéraud studied 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.

The proposed approach combines the use of asymptotic complexity functions provided by the programmer, with runtime measurements to estimate the constant factors that apply. Exploiting these two sources of information makes it possible 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 [52], 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 [14].

7.8. Program Verification and Formal Languages

Participant: Arthur Charguéraud.

- Armaël Guéneau, a PhD student advised by A. Charguéraud and F. Pottier (Cambium), has developed a formal proof of the functional correctness and the asymptotic complexity of a state-of-the-art incremental cycle detection algorithm due to Bender, Fineman, Gilbert, and Tarjan. This work moreover proposes a simple change that allows the algorithm to be regarded as genuinely online. The verification proof is carried out by exploiting Separation Logic with Time Credits, in the CFML tool, to simultaneously verify the correctness and the worst-case amortized asymptotic complexity of the modified algorithm. This work was published at ITP'19 [17]. It leverages previous work on the extension of the CFML verification tool to allow the specification of the asymptotic complexity of higher-order, imperative programs [55], and shows that this framework scales up to larger, more complex programs.
- Arthur Charguéraud, together with Jean-Christophe Filliâtre and Cláudio Lourenço (CNRS, Inria and Université Paris Saclay), and Mário Pereira (NOVA LINCS & DI, Universidade Nova de Lisboa), developed a behavioral specification language for OCaml, called GOSPEL. It is designed to enable modular verification of data structures and algorithms. Compared with writing specifications directly in Separation Logic, it provides a high-level syntax that greatly improves conciseness and makes it accessible to programmers with no familiarity with Separation Logic. GOSPEL is applied to the development of a formally verified library of general-purpose OCaml data structures. This work was published at the World Congress on Formal Methods (FM) 2019 [15].

7.9. Improvement of Schnaps on multi-GPU nodes using the LAHeteroprio Scheduler

Participant: Bérenger Bramas.

The TONUS team has developed Schnaps, a discontinuous finite element solver with OpenCL and StarPU. The team members have been facing challenges in the scalability of their application when using more than one GPU. This has been the starting point of a collaboration in which Bérenger Bramas has participated in the development of Schnaps and plugged its StarPU scheduler called LAHeteroprio [9]. The improvements obtained were significant and included in a paper [50] (currently under revision).

The potential of LAHeteroprio is now demonstrated. However, setting up this scheduler remains a complicated task. Therefore, we plan to work on its automatic configuration, which will require us to perform on the fly analysis of the graph of tasks.

7.10. Improving Parallel Executions by Increasing Task Granularity in Task-based Runtime Systems using Acyclic DAG Clustering

Participants: Bérenger Bramas, Alain Ketterlin.

Bérenger Bramas and Alain Ketterlin collaborate with the TONUS team in the development of a parallel solver for the resolution of conservative hyperbolic upwind kinetic of unstructured tokamaks [49]. In their methods, they must solve the transport equation on an unstructured mesh, which can be seen as having a wave propagating from neighbor-to-neighbor. The resulting computation can be represented using a direct acyclic graph (DAG) of operations, where each operation is a tiny task. Therefore, Bérenger Bramas and Alain Ketterlin contributed mainly on two aspects. First, they have proposed a highly optimized lock-free parallel implementation of the solution based on atomic instructions. Second, they have improved an existing algorithm from the literature to cluster a DAG of tasks with the aim of increasing the granularity of the tasks and to reduce the overhead of the parallelization consequently. This new approach has been accepted in a dedicated paper (accepted but not yet published).

7.11. FMM Kernel for the Integral Equation Formulation of the N-body Dielectric Spheres Problem

Participant: Bérenger Bramas.

Bérenger Bramas worked with Benjamin Stamm and Muhammad Hassan (RWTH) to create a kernel for the fast multipole method (FMM). The kernel relies on the previously developed kernel with spherical harmonics and accelerated by rotations. It has been extended to accept spherical harmonics (with orders different from the ones used in the kernel) instead of points as input. The kernel allowed us to accelerate the computation and was used for a complexity analysis that has been submitted [54].

7.12. Automatic Task-Based Parallelization using Source to Source Transformations

Participants: Bérenger Bramas, Garip Kusolgu.

Bérenger Bramas and Garip Kusolgu worked on a new approach to parallelize automatically any application written in an object-oriented language. The main idea is to parallelize a code as an HPC expert would do it using the task-based method. With this aim, they created a new source-to-source compiler on top of CLang-LLVM called APAC. APAC is able to insert tasks in a source-code by evaluating data accesses and thus generating the correct dependencies. An important and challenging part of the work consists in managing the granularity, which requires to work both statically on the code but also by delegating decisions at runtime.

7.13. Large Scale Particle Fusion Algorithm for Tracing Systems in Fluid Mechanics Applications

Participant: Bérenger Bramas.

Bérenger Bramas worked with Michael Wilczek and Cristian Lalescu (Max Planck Institute for Dynamics and Self-Organization) in designing a new method to merge particles in a large scale application (*i.e.,* designed to run on thousands of computing nodes). In this context, the particles are originally used in a tracing system to extract information from a vector field in fluid mechanics. However, the physicists are now interested having the particles interacting and even fusioning. Due to the constraints of large scale computing, the system tries to reduce the number and amount of communications. This development has been done in the TurTLE application (not publicly available) and is currently under evaluation.

7.14. Pipelined Multithreaded Code Generation

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

State-of-the-art automatic polyhedral parallelizers extract and express parallelism as isolated parallel loops. For example, the Pluto high-level compiler generates and annotates loops with `#pragma omp parallel for` directives. In this work, we took advantage of pipelined multithreading, a parallelization strategy that can address a wider class of codes, currently not handled by automatic parallelizers. Pipelined multithreading requires interlacing iterations of some loops in a controlled way that enables the parallel execution of these iterations.

This work has been accepted for presentation at the International Workshop on Polyhedral Compilation Techniques (IMPACT 2020), in conjunction with HiPEAC '20 (Jan. 2020, Bologna, Italy).

7.15. Raster Image Processing (RIP) Optimization

Participants: Cédric Bastoul, Paul Godard, Vincent Loechner.

In the context of our collaboration with the Caldera company, we are interested in original challenges for the computer systems in charge of driving very wide printer farms and very fast digital presses.

We explored new approaches inspired by the high performance computing field to speedup the graphics processing (RIP) necessary to digital printing. To achieve this goal, we developed a distributed system which provides the adequate flexibility and performance by exploiting and optimizing both processing and synchronization techniques. Our architecture meets the specific constraints on generating streams for printing purpose. We performed an evaluation of our solution and provided experimental evidence of its great performance and viability. This work has been presented at the 2019 IEEE International Parallel and Distributed Processing Symposium Workshop (IPDPSW): PDSEC '19, in May 2019, Rio de Janeiro.

The second topic we worked on during this collaboration is an out-of-core and out-of-place rectangular matrix transposition and rotation algorithm. An originality of our processing algorithm is to rely on an optimized use of the page cache mechanism. It is parallel, optimized by several levels of tiling and independent of any disk block size. We evaluated our approach on four common storage configurations: HDD, hybrid HDD-SSD, SSD and software RAID 0 of several SSDs. We showed that it brings significant performance improvement over a hand-tuned optimized reference implementation developed by the Caldera company and we confront it against the roofline speed of a straight file copy. This work is under submission in the IEEE Transaction on Computers.

Paul Godard has defended his PhD thesis on Dec. 16th, 2019.

7.16. Static Versus Dynamic Memory Allocation

Participant: Vincent Loechner.

Vincent Loechner and Toufik Baroudi (PhD student, Univ. Batna, Algeria) compared the performance of linear algebra kernels using different array allocation modes: as static declared arrays or as dynamically allocated arrays of pointers. They studied the possible reasons of the difference in performance of parallelized or sequential linear algebra kernels on two different architectures: an AMD (Magny-Cours) and an Intel Xeon (Haswell-EP). Static or dynamic memory allocation has an impact on performance in many cases. Both the processor architecture and the compiler can provoke significant and sometimes surprising variations in the number of cache misses and vectorization opportunities taken by the compiler.

This work has been accepted for presentation at the International Workshop on Polyhedral Compilation Techniques (IMPACT 2020), in conjunction with HiPEAC '20 (Jan. 2020, Bologna, Italy).

7.17. Automatic Adaptive Approximation for Stencil Computations

Participants: Maxime Schmitt, Cédric Bastoul.

This work has been done in collaboration with Philippe Helluy (TONUS).

Approximate computing is necessary to meet deadlines in some compute-intensive applications like simulation. Building them requires a high level of expertise from the application designers as well as a significant development effort. Some application programming interfaces greatly facilitate their conception but they still heavily rely on the developer's domain-specific knowledge and require many modifications to successfully generate an approximate version of the program. In this work we designed new techniques to semi-automatically discover relevant approximate computing parameters. We believe that superior compiler-user interaction is the key to improved productivity. After pinpointing the region of interest to optimize, the developer is guided by the compiler in making the best implementation choices. Static analysis and runtime monitoring are used to infer approximation parameter values for the application. We evaluated these techniques on multiple application kernels that support approximation and show that with the help of our method, we achieve similar performance as non-assisted, hand-tuned version while requiring minimal intervention from the user.

These techiques and the underlying compiler infrastructure are a significant output of collaboration with the 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 CC international conference [18].

Maxime Schmitt has defended his PhD thesis on Sep. 30th, 2019 [8].

CASH Project-Team

6. New Results

6.1. Dataflow-explicit futures

Participants: Ludovic Henrio, Matthieu Moy, Amaury Maillé.

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 [43], 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:

- With collaborators from University of Uppsala and University of Oslo we worked on the design of programming constructs mixing implicit and dataflow-explicit futures (DeF). This has been published in ECOOP 2019 [10].
- Amaury Maillé did his internship in the Cash team (advised by Matthieu Moy and Ludovic Henrio), he worked on an implementation of DeF in the Encore language. This raised a difficulty regarding the interaction of DeF with generic types that has been partially solved. Now we need to generalize our approach to completely solve the issue.

6.2. Distributed futures

Participant: Ludovic Henrio.

We proposed the definition of *distributed futures*, a construct that provides at the same time a data container similar to a distributed vector, and a single synchronization entity that behaves similarly to a standard future. This simple construct makes it easy to program a composition, in a task-parallel way, of several massively data-parallel tasks. This work will be presented in Sac 2020 (we are currently working on the final version of the paper). This work is realised in collaboration with Pierre Leca and Wijnand Suijlen (Huawei Technologies), and Françoise Baude (Université Côte d'Azur, CNRS, I3S).

6.3. 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. After 2 meetings in 2019, A journal article is still being written but the visit of Reiner Hähnle in the Cash team during two months (as invited professor) in Spring 2019 should allow us to make faster progress on the topic.

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).

6.4. 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:

- we extended the work to support the manipulation of overlapping array. This was accepted as an extended version of our conference paper (presented at 4PAD 2018). It will be published in the JLAMP journal in 2020 [3].

6.5. 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. This work is realized with Eric Madelaine (Inria Sophia-Antipolis) and Rabéa Ameur Boulifa (Telecom ParisTech). A journal article has been written and will be submitted in January 2020.
- A translation from BIP model to open pNets has being formalized and encoded, this work is done in collaboration with Simon Bliudze (Inria Lille). More precisely, we extend the theory of architectures developed previously for the BIP framework with the elements necessary for handling data: definition and operations on data domains, syntax and semantics of composition operators involving data transfer. To verify that individual architectures do enforce their associated properties , we provide an encoding into open pNets, an intermediate model that supports SMT-based verification. This work has been published in Coordination 2019 [6].

These works are under progress and should be continued in 2020.

6.6. 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 have been published in TACAS 2019 [11].

6.7. A Survey on Verified Reconfiguration

Participant: Ludovic Henrio.

We are conducting a survey on the use of formal methods to ensure safety of reconfiguration of distributed system, that is to say the runtime adaptation of a deployed distributed software system. The survey article is written together with Hélène Coullon and Simon Robillard (IMT Atlantique, Inria, LS2N, UBL), and Frédéric Loulergue (Northern Arizona University). Hélène Coullon is the coordinator and we expect the article to be submitted in 2020.

6.8. A Survey on Parallelism and Determinacy

Participants: Ludovic Henrio, Laure Gonnord, Matthieu Moy, Christophe Alias.

We have started to investigate on the solutions that exist to ensure complete or partial determinacy in parallel programs. The objective of this work is to provide a survey based on the different kinds of solutions that exist to ensure determinism or at least limit data-races in concurrent execution of programs. The study will cover language-based, compilation-based and also runtime-based solutions. We started the bibliographic studies in 2019. The objective of this work is to write and submit a survey article in 2020.

This work, coordinated by Laure Gonnord and Ludovic Henrio, also involves contributors outside the CASH team. For the moment Gabriel Radanne (Inria Paris) and Lionel Morel (CEA).

6.9. Pipeline-aware Scheduling of Polyhedral Process Networks

Participants: Christophe Alias, Julien Rudeau.

The polyhedral model is a well known framework to develop accurate and optimal automatic parallelizers for high-performance computing kernels. It is progressively migrating to high-level synthesis through polyhedral process networks (PPN), a dataflow model of computation which serves as intermediate representation for high-level synthesis. Many locks must be overcome before having a fully working polyhedral HLS tool, both from a front-end (C → PPN) and back-end (PPN → FPGA) perspective. In this work [15], we propose a front-end scheduling algorithm which reorganizes the computation of processes to maximize the pipeline efficiency of the processes' arithmetic operators. We show that our approach improve significantly the overall latency as well as the pipeline efficiency.

6.10. A Compiler Algorithm to Guide Runtime Scheduling

Participants: Christophe Alias, Samuel Thibault, Laure Gonnord.

Task-level parallelism is usually exploited by a runtime scheduler, after tasks are mapped to processing units by a compiler. In this report, we propose a compilation-centric runtime scheduling strategy. We propose a complete compilation algorithm to split the tasks in three parts, whose properties are intended to help the scheduler to take the right decisions [16]. In particular, we show how the polyhedral model may provide a precious help to compute tricky scheduling and parallelism informations. Our compiler is available and may be tried online at http://foobar.ens-lyon.fr/kut.

This is a joint work with University of Bordeaux, which will be continued next year.

6.11. fkcc: the Farkas Calculator

Participant: Christophe Alias.

We propose a new domain-specific language and a tool, FKCC, to prototype program analyses and transformations exploiting the affine form of Farkas lemma. Our language is general enough to prototype in a few lines sophisticated termination and scheduling algorithms. The tool is freely available and may be tried online via a web interface. We believe that FKCC is the missing chain to accelerate the development of program analyses and transformations exploiting the affine form of Farkas lemma.

This work has been presented in the TAPAS'19 workshop [13] and will be presented at the IMPACT'20 workshop [13].

6.12. Standard-compliant Parallel SystemC simulation of Loosely-Timed Transaction Level Models

Participant: Matthieu Moy.

To face the growing complexity of System-on-Chips (SoCs) and their tight time-tomarket constraints, Virtual Prototyping (VP) tools based on SystemC/TLM must get faster while keeping accuracy. However, the Accellera SystemC reference implementation remains sequential and cannot leverage the multiple cores of modern workstations. In this paper, we present a new implementation of a parallel and standard-compliant SystemC kernel, reaching unprecedented performances. By coupling a parallel SystemC kernel and memory access monitoring, we are able to keep SystemC atomic thread evaluation while leveraging the available host cores. Evaluations show a ×19 speed-up compared to the Accellera SystemC kernel using 33 host cores reaching speeds above 2000 Million simulated Instructions Per Second (MIPS).

This work will be published at the ASP-DAC 2020 conference.

6.13. Response time analysis of dataflow applications on a many-core processor with shared-memory and network-on-chip

Participant: Matthieu Moy.

We consider hard real-time applications running on many-core processor containing several clusters of cores linked by a Network-on-Chip (NoC). Communications are done via shared memory within a cluster and through the NoC for inter-cluster communication. We adopt the time-triggered paradigm, which is well-suited for hard real-time applications, and we consider data-flow applications, where communications are explicit.

We extend the AER (Acquisition/Execution/Restitution) execution model to account for all delays and interferences linked to communications, including the interference between the NoC interface and the memory. Indeed, for NoC communications, data is first read from the initiator's local memory, then sent over the NoC, and finally written to the local memory of the target cluster. Read and write accesses to transfer data between local memories may interfere with shared-memory communication inside a cluster, and, as far as we know, previous work did not take these interferences into account.

Building on previous work on deterministic network calculus and shared memory interference analysis, our method computes a static, time-triggered schedule for an application mapped on several clusters. This schedule guarantees that deadlines are met, and therefore provides a safe upper bound to the global worst-case response time.

This work was published at RTNS 2019 [14].

6.14. Smart placement of dynamically allocated objects for heterogeneous memory

Participant: Matthieu Moy.

As part of a partnership with the CITI laboratory (Tristan Delizy's PhD, co-supervised with Guillaume Salagnac and Tanguy Risset), we worked on dynamic memory memory allocation for embedded systems with heterogeneous memory. Unlike cache-based systems, our target architecture exposes several memory banks with different performance characteristics directly to the software, without any hardware mechanism like a cache or an MMU for memory management. The software needs to chose which memory bank to use at allocation time, and cannot change this choice afterwards. We proposed a profiling-based placement policy that is shown to be near-optimal for several applications, and performs much better than naive placement policies especially for systems with a small fraction of fast memory.

This work documented as part of Tristan Delizy's Ph.D manuscript, and we plan to submit it for a journal publication in 2020.

6.15. 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 in 2018 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 presented to VMCAI'19 on Model Checking and Abstract Interpretation [5] and has received the best paper award of the conference.

6.16. 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), and rephrase a complete value and validity

In 2019, the formalisation has been rephrased in a simplier way and extented to deal with more llvm-related semantics (undefined behavior, poisoned values) and we started a collaboration with David Monniaux, Verimag, on this topic.

The paper is being rewritten and we are also writing a project on which we would extend our method to mode complex polyhedral transformations in a context of formally verified tools.

6.17. Decision results for solving Horn Clauses with arrays

Participants: Laure Gonnord, Julien Braine.

Many approaches exist for verifying programs operating on Boolean and integer values (e.g. abstract interpretation, counterexample-guided abstraction refinement using interpolants), but transposing them to array properties has been fraught with difficulties. In the context of the Phd of Julien Braine, we propose to work directly on horn clauses, because we think that it is a suitable intermediate representation for verifying programs.

Currently, two techniques strike out to infer very precise quantified invariants on arrays using Horn clauses: a quantifier instantiation method [1] and a cell abstraction method that can be rephrased on Horn clauses. However, the quantifier instantiation method is parametrized by an heuristic and finding a good heuristic is a major challenge, and the cell abstraction method uses an abstract interpretation to completely remove arrays and is limited to linear Horn clauses. We combine these two techniques. We provide an heuristic for the quantifier instantiation method of [29] by using the ideas from the cell abstraction method of [48] and discover a requirement such that, when met, the heuristic is complete, that is, there is no loss of information by using that heuristic. Furthermore, we prove that Horn clauses that come from program semantic translation verify the requirement and therefore, we have an optimal instantiation technique for program analysis.

This work is done in collaboration with David Monniaux (Verimag), coadvisor of the PhD of Julien Braine. A journal paper is currently being written for submission early 2020.

6.18. Scheduling Trees

Participants: Laure Gonnord, Paul Iannetta.

As a first step to schedule non polyhedral computation kernels, we investigated the tree datastructure. A large bibliography on tree algorithmics and complexity leds us to chose to work on balanced binary trees, for which we have designed algorithms to change their memory layout into adjacent arrays. We rephrased the classical algorithms (construction, search, destruction ...) in this setting, and implemented them in C.

The conclusion of this study is unfortunately negative : the locality gain in transforming trees into linear structures is not contrabalanced by a better cache usage, all our codes have been slowed down in the process. Our experiments are still in progress, but our hypothesis is that our trees are too sparse to be more clever that the *malloc* implementation.

A research paper will be published early 2020. This work is done in collaboration with Lionel Morel (CEA Grenoble), coadvisor of the PhD of Paul Iannetta.

6.19. Formalisation of the Polyhedral Model

Participants: Laure Gonnord, Paul Iannetta.

Last year, together with Lionel Morel (Insa/CEA) and Tomofumi Yuki (Inria, Rennes), we revisited the polyhedral model's key analysis, dependency analysis, published in a research report [44]. This year we pursued in this direction. We have now a better formalisation, and a better understanding of the expressivity and applicability.

We still have one step to study in order to be able to have a full semantic polyhedral model: properly formalise code scheduling and code generation within our semantic model.

This work is made in collaboration with Lionel Morel (CEA Grenoble) who coadvise Paul Iannetta.

6.20. Semantics diffs in LLVM

Participants: Laure Gonnord, Matthieu Moy.

Laure Gonnord and Matthieu Moy have coadvised a Master research Project ("TER") early in 2019 , whose objective was to study the LLVM LLVM compiler infrastructure with software engineering techniques in order to characterise how sequences of code analyses and transformations behave. The project has lead to a sequence of tools to evaluate experimentally how a sequence of passes influence performance.

Laure Gonnord and Matthieu Moy have, together with Sebastien Mosser, coadvised a second internship at UQAM for three months, between May and July 2019. During his internship, Sebastien Michelland has demonstrated that textual diffs are not sufficient to fully characterise the behaviours of code transformation inside compilers. He analysed llvm-diff, a tool of the distribution that makes an analysis at the intermediate representation level, and gives first hints to define a proper notion of semantic diff for this application.

For these interships two research reports have been produced.

This work was done in the context of an ongoing collaboration with Sebastien Mosser, previously in Nice, and now at UQAM. An Inria associate team was proposed for 2020-2023 on similar topics.

CORSE Project-Team

6. New Results

6.1. Compiler Optimizations and Analysis

Participants: Fabrice Rastello, Manuel Selva, Fabian Grüber, Diogo Sampaio [CORSE, Inria], Christophe Guillon [STMicroelectronics], P. Sadayappan [OSU, USA], Louis-Noël Pouchet [CSU, USA], Atanas Rountev [OSU, USA], Richard Veras [LSU, USA], Rui Li [UoU, USA], Aravind Sukumaran-Rajam [OSU, USA], Tse Meng Low [CMU, USA].

Our current efforts with regard to code optimization follows two directions. 1. The first consists in improving compiler optimization techniques by considering pattern specific applications such as those related to machine learning. Our first result presented at SC 2019 [10] focuses on tensor contractions. 2. The second consists in developing dynamic analysis based performance debugging tools. Our first results published at PPoPP 2019 [9] and TACO 2020 [7] shows a scalable approach that compresses an execution trace obtained from binary instrumentation and analyses it using a polyhedral compiler.

6.1.1. Analytical Cache Modeling and Tilesize Optimization for Tensor Contractions

Data movement between processor and memory hierarchy is a fundamental bottleneck that limits the performance of many applications on modern computer architectures. Tiling and loop permutation are key techniques for improving data locality. However, selecting effective tile-sizes and loop permutations is particularly challenging for tensor contractions due to the large number of loops. Even state-of-the-art compilers usually produce sub-optimal tile-sizes and loop permutations, as they rely on naïve cost models. In this work we provide an analytical model based approach to multilevel tile size optimization and permutation selection for tensor contractions. Our experimental results show that this approach achieves comparable or better performance than state-of-the-art frameworks and libraries for tensor contractions.

This work is the fruit of the collaboration 8.3.1.1 with OSU. It has been presented at ACM/IEEE International Conference for High Performance Computing, Networking, Storage, and Analysis, SC 2019 [10].

6.1.2. Profiling-based Polyhedral Optimization Feedback

This work addresses the problem of reconstructing a compact (static) representation of a binary execution, automatically detecting hot regions and enabling precise feedback about optimization opportunities potentially missed by the compiler. Our framework handles codes with irregular accesses, pointers with indirections, inter-procedural or recursive loop regions. By enabling binary execution analysis we are able to discover runtime properties (i.e., the ability to form a compact representation) as well as inter-procedural optimization opportunities that cannot be uncovered by standard static analyses. Our design choices were driven towards achieving portability, both in terms of targeted architecture, but also in terms of programming environment (e.g., being robust to arbitrary programming language, compiler, use of third-party binaries, etc.).

A compact and yet precise inter-procedural dynamic dependence graph (DDG) is first computed via: 1. a new instrumentation framework based on QEMU; 2. the use of a new concept of inter-procedural loop-nesting tree; 3. followed by new techniques we introduce for folding, clamping, and widening of the DDG to agglomerate dynamic dependence instances into polyhedra of integer points whenever possible. State-of-the-art polyhedral analysis and transformation systems we specifically modified to provide useful feedback to the user is then used. We extensively evaluate our tool on numerous benchmarks, demonstrating the pratical usefulness of our tool-chain.

This work is the fruit of the collaboration 8.3.1.1 with OSU and and the past collaboration Nano2017 with STMicroelectronics. The main contributions has been presented at the ACM conference on Principles and Practice of Parallel Programming, PPoPP 2019 [9]. The new techniques that allow to build the polyhedral representation from the instrumented execution in a scalable way lead to a separate publication in the ACM Transactions on Architecture and Code Optimization, TACO 2020 [7].

6.2. Extraction of Periodic Patterns of Scientific Applications to Identify DVFS Opportunities

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 developed a tool suite resorting to fine-grain profiling and periodicity analysis to identify optimization opportunities for both performance and power-efficiency. It leverages the fact that a large share of HPC parallel applications are constituted of a restrained set of compute kernels executed a huge number of times to extract periodic patterns representative of the aforementioned kernels. By doing so, our tool offers a simple and condensed proxy to analyze and predict the behavior of complex parallel applications. For instance, we were able to identify and extract periodic patterns for a panel of reference HPC applications such as NAMD and NEMO. Then, as an example of the many ways to exploit the aforementioned extracted periodic patterns, we evaluated the impact of the CPU frequency on the latter. As a result, we were able to identify DVFS opportunities we plan to exploit in a future work.

6.3. Runtime Monitoring, Verification, and Enforcement

Participants: Antoine El-Hokayem [Univ. Grenoble Alpes, Verimag], Yliès Falcone, Thierry Jéron [Inria Rennes], Ali Kassem, Hervé Marchand [Inria Rennes], Srinivas Pinisetty [IIT Bhubaneswar], Matthieu Renard [Foxi], Antoine Rollet [Université de Bordeaux], César Sànchez [IMDEA Madrid], Gerardo Schneider [University of Gothenborg].

Our contributions in the domain of runtime monitoring, verification, and enforcement are threefold. First, we contributed to the publication of general papers aimed to structure the community by publishing a tutorial on runtime enforcement of timed properties [16], a review of the first five years of the competition on runtime verification [15] and a survey of future challenges of runtime verification [6]. We also concluded some other previous work by realizing journal publications on the topics of decentralized runtime verification [3] and on runtime enforcement of timed properties [5]. We started a new activity on monitoring for security properties, and more particularly on the detection of fault-injection attacks [12].

6.3.1. On the Runtime Enforcement of Timed Properties

This work [16] is concerned with runtime enforcement which refers to the theories, techniques, and tools for enforcing correct behavior of systems at runtime. We are interested in such behaviors described by specifications that feature timing constraints formalized in what is generally referred to as timed properties. This tutorial presents a gentle introduction to runtime enforcement (of timed properties). First, we present a taxonomy of the main principles and concepts involved in runtime enforcement. Then, we give a brief overview of a line of research on theoretical runtime enforcement where timed properties are described by timed automata and feature uncontrollable events. Then, we mention some tools capable of runtime enforcement, and we present the TiPEX tool dedicated to timed properties. Finally, we present some open challenges and avenues for future work.

6.3.2. Detecting Fault Injection Attacks with Runtime Verification

This work [12] is concerned with fault injections which are increasingly used to attack/test secure applications. In this paper, we define formal models of runtime monitors that can detect fault injections that result in test inversion attacks and arbitrary jumps in the control flow. Runtime verification monitors offer several advantages. The code implementing a monitor is small compared to the entire application code. Monitors have a formal semantics; and we prove that they effectively detect attacks. Each monitor is a module dedicated to detecting an attack and can be deployed as needed to secure the application. A monitor can run separately from the application or it can be weaved inside the application. Our monitors have been validated by detecting simulated attacks on a program that verifies a user PIN.

6.3.3. International Competition on Runtime Verification (CRV)

In this work [15], we review the first five years of the international Competition on Runtime Verification (CRV), which began in 2014. Runtime verification focuses on verifying system executions directly and is a useful lightweight technique to complement static verification techniques. The competition has gone through a number of changes since its introduction, which we highlight in this paper.

6.3.4. A Survey of Challenges for Runtime Verification from Advanced Application Domains (beyond software)

In this work [6], we survey the future challenges for runtime verification. Typically, the two main activities in runtime verification efforts are the process of creating monitors from specifications, and the algorithms for the evaluation of traces against the generated monitors. Other activities involve the instrumentation of the system to generate the trace and the communication between the system under analysis and the monitor. Most of the applications in runtime verification have been focused on the dynamic analysis of software, even though there are many more potential applications to other computational devices and target systems. In this paper we present a collection of challenges for runtime verification extracted from concrete application domains, focusing on the difficulties that must be overcome to tackle these specific challenges. The computational models that characterize these domains require to devise new techniques beyond the current state of the art in runtime verification.

6.3.5. On the Monitoring of Decentralized Specifications Semantics, Properties, Analysis, and Simulation

In this work [3], we define two complementary approaches to monitor decentralized systems. The first relies on those with a centralized specification, i.e, when the specification is written for the behavior of the entire system. To do so, our approach introduces a data-structure that i) keeps track of the execution of an automaton, ii) has predictable parameters and size, and iii) guarantees strong eventual consistency. The second approach defines decentralized specifications wherein multiple specifications are provided for separate parts of the system. We study two properties of decentralized specifications pertaining to monitorability and compatibility between specification and architecture. We also present a general algorithm for monitoring decentralized specifications. We map three existing algorithms to our approaches and provide a framework for analyzing their behavior. Furthermore, we introduce THEMIS, a framework for designing such decentralized algorithms and simulating their behavior. We show the usage of THEMIS to compare multiple algorithms and verify the trends predicted by the analysis by studying two scenarios: a synthetic benchmark and a real example.

6.3.6. Optimal Enforcement of (timed) Properties with Uncontrollable Events

This work deals with runtime enforcement of untimed and timed properties with uncontrollable events [5]. Runtime enforcement consists in defining and using mechanisms that modify the executions of a running system to ensure their correctness with respect to a desired property. We introduce a framework that takes as input any regular (timed) property described by a deterministic automaton over an alphabet of events, with some of these events being uncontrollable. An uncontrollable event cannot be delayed nor intercepted by an enforcement mechanism. Enforcement mechanisms should satisfy important properties, namely soundness, compliance and optimality – meaning that enforcement mechanisms should output as soon as possible correct executions that are as close as possible to the input execution. We define the conditions for a property to be enforceable with uncontrollable events. Moreover, we synthesise sound, compliant and optimal descriptions of runtime enforcement mechanisms at two levels of abstraction to facilitate their design and implementation.

6.4. Teaching of Algorithms, Programming, Debugging, and Automata

Participants: Florent Bouchez Tichadou, Yliès Falcone, Théo Barollet, Antoine Clavel, Thomas Hervé, Anthony Martinez, Beryl Piasentin, Steven Sengchanh.

This domain is a new axis of the Corse team. Our goal here is to combine our expertise in compilation and teaching to help teachers and learners in computer science fields such as programming, algorithms, data strucures, automata, or more generally computing litteracy. The most important project in this regard is the automated generation and recommendation of exercises using artificial intelligence, a thesis that started this year. Other projects focus on tools to help learning through visualization (data structures, debugger, automata) or gamification (AppoLab), and are the source of many internships that give younger students experience in a research team.

6.4.1. AI4HI: Artificial Intelligence for Human Intelligence

In an ideal educative world, each learner would have access to individual pedagogical help, tailored to its needs. For instance, a tutor who could rapidly react to the questions, and propose pedagogical contents that match the learner's kills, and who could identify and work on his or her weaknesses. However, the real world imposes constraints that make this individual pedagogical help hard to achieve.

The goal of the AI4HI project is to combine the new advances in artificial intelligence with the team's skills in compilation and teaching to aid teaching through the automated generation and recommendation of exercises to learners. In particular, we target the teaching of programming and debugging to novices. This system will propose exercises that match the learners' needs and hence improve the learning, progression, and self-confidence of learners.

This projet has received an "Action Exploratoire" funding from Inria and Théo Barollet started his PhD this September so is still in its early stages.

6.4.2. AppoLab

Classical teaching of algorithms and low-level data structures is often tedious and unappealing to students. AppoLab is an online platform to engage students in their learning by including gamification in Problem-Based Learning. In its core, it is a server with scripted "exercises". Students can communicate with the server manually, but ultimately they 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.

6.4.3. Data Structures and Program Visualization at Runtime

Debuggers are powerful tools to observe a program behaviour and find bugs but they have a hard learning curve. 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 developed a prototype, Moly, a GDB extension that explores a program runtime memory and analyze its data structures. It provides an interface with an external visualizer, Lotos, through a formatted output. Work has also started to include a tutorial on how to use GDB and these extensions.

6.4.4. Aude

Aude is a pedagogical software for manipulating, learning, and teaching finite state automata and the automata theory. It is used by the students in the second year of the bachelor in computer science at Univ. Grenoble Alpes. It allows students to get acquainted and autonomously work on the concepts involved in the theory of regular languages and automata.

PACAP Project-Team

7. New Results

7.1. Compilation and Optimization

Participants: Loïc Besnard, Caroline Collange, Byron Hawkins, Erven Rohou, Bahram Yarahmadi.

7.1.1. Optimization in the Presence of NVRAM

Participants: Erven Rohou, Bahram Yarahmadi.

A large and increasing number of Internet-of-Things devices are not equipped with batteries and harvest energy from their environment. Many of them cannot be physically accessed once they are deployed (embedded in civil engineering structures, sent in the atmosphere or deep in the oceans). When they run out of energy, they stop executing and wait until the energy level reaches a threshold. Programming such devices is challenging in terms of ensuring memory consistency and guaranteeing forward progress.

7.1.1.1. Checkpoint Placement based Worst-Case Energy Consumption

Previous work has proposed to insert checkpoints in the program so that execution can resume from well-defined locations. We propose to define these checkpoint locations based on worst-case energy consumption of code sections, with limited additional effort for programmers. As our method is based upon worst-case energy consumption, we can guarantee memory consistency and forward progress.

This work has been presented at the Compas 2019 conference.

7.1.1.2. Dynamic Adaptive Checkpoint Placement

Previous work has proposed to back-up the volatile states which are necessary for resuming the program execution after power failures. They either do it at compile time by placing checkpoints into the control flow of the program or at runtime by leveraging voltage monitoring facilities and interrupts, so that execution can resume from well-defined locations after power failures. We propose for the first time a dynamic checkpoint placement strategy which delays checkpoint placement and specialization to the runtime and takes decisions based on the past power failures and execution paths that are taken. We evaluate our work on a TI MSP430 device, with different types of benchmarks as well as different uninterrupted intervals, and we measure the execution time. We show that our work can outperform compiler-based state-of-the-art with memory footprint kept under the control.

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

7.1.2. Dynamic Binary Optimization

Participant: Erven Rohou.

7.1.2.1. Guided just-in-time specialization

JavaScript's portability across a vast ecosystem of browsers makes it today a core building block of the web. Yet, building efficient systems in JavaScript is still challenging. Because this language is so dynamic, JavaScript programs provide little information that just-in-time compilers can use to carry out safe optimizations. Motivated by this observation, we propose to guide the JIT compiler in the task of code specialization. To this end, we have augmented [17] the language with an annotation that indicates which function call sites are likely to benefit from specialization. To support the automatic annotation of programs, we have introduced a novel static analysis that identifies profitable specialization points. We have implemented our ideas in JavaScriptCore, the built-in JavaScript engine for WebKit. The addition of guided specialization to this engine required us to change it in several non-trivial ways. Such changes let us observe speedups of up to 1.7× on programs present in synthetic benchmarks.

7.1.2.2. Run-time parallelization and de-parallelization

Runtime compilation has opportunities to parallelize code which are generally not available using static parallelization approaches. However, the parallelized code can possibly slowdown the performance due to unforeseen parallel overheads such as synchronization and speculation support pertaining to the chosen parallelization strategy and the underlying parallel platform. Moreover, with the wide usage of heterogeneous architectures, such choice options become more pronounced. We consider [22] an adaptive form of the parallelization operation, for the first time. We propose a method for performing on-stack de-parallelization for a parallelized binary loop at runtime, thereby allowing for rapid loop replacement with a more optimized one. We consider a loop parallelization strategy and propose a corresponding de-parallelization method. The method relies on stopping the execution at safe points, gathering threads' states, producing a corresponding serial code, and continuing execution serially. The decision to de-parallelize or not is taken based on the anticipated speedup. To assess the extent of our approach, we have conducted an initial study on a small set of programs with various parallelization overheads. Results show up to 4× performance improvement for a synchronization intense program on a 4-core Intel processor.

With the multicore trend, the need for automatic parallelization is more pronounced, especially for legacy and proprietary code where no source code is available and/or the code is already running and restarting is not an option. We engineer [21] a mechanism for transforming at runtime a frequent for-loop with no data dependencies in a binary program into a parallel loop, using on-stack replacement. With our mechanism, there is no need for source code, debugging information or restarting the program. Also, the mechanism needs no static instrumentation or information. The mechanism is implemented using the Padrone binary modification system and `pthreads`, where the remaining iterations of the loop are executed in parallel. The mechanism keeps the running program state by extracting the targeted loop into a separate function and copying the current stack frame into the corresponding frames of the created threads. Initial study is conducted on a set of kernels from the Polybench workload. Experimental results show from 2× to 3.5× speedup from sequential to parallelized code on four cores, which is similar to source code level parallelization.

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

7.1.3. Automatic and Parametrizable Memoization

Participants: Loïc Besnard, Erven Rohou.

Improving execution time and energy efficiency is needed for many applications and usually requires sophisticated code transformations and compiler optimizations. One of the optimization techniques is memoization, which saves the results of computations so that future computations with the same inputs can be avoided. We propose [16] a framework that automatically applies memoization techniques to C/C++ applications. The framework is based on automatic code transformations using a source-to-source compiler and on a memoization library. With the framework users can select functions to memoize as long as they obey to certain restrictions imposed by our current memoization library. We show the use of the framework and associated memoization technique and the impact on reducing the execution time and energy consumption of four representative benchmarks. 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).

7.1.4. Autotuning

Participants: Loïc Besnard, Erven Rohou.

The ANTAREX FET HPC project relies on a Domain Specific Language (DSL) based on Aspect Oriented Programming (AOP) concepts to allow applications to enforce extra functional properties such as energy-efficiency and performance and to optimize Quality of Service (QoS) 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. We present [20] an overview of the key outcome of the project, the ANTAREX DSL, and some of its capabilities through a number of examples, including how the DSL is applied in the context of the project use cases. We demonstrated [30] tools and techniques in two domains: computational drug discovery, and online vehicle navigation.

7.1.5. 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 was partially done within the context of the ANTAREX FET HPC collaborative project, collaboration is currently ongoing with University of Porto, Portugal.

7.1.6. Hardware/Software JIT Compiler

Participant: Erven Rohou.

Single-ISA heterogeneous systems (such as ARM big.LITTLE) are an attractive solution for embedded platforms as they expose performance/energy trade-offs directly to the operating system. Recent works have demonstrated the ability to increase their efficiency by using VLIW cores, supported through Dynamic Binary Translation (DBT) to maintain the illusion of a single-ISA system. However, VLIW cores cannot rival with Out-of-Order (OoO) cores when it comes to performance, mainly because they do not use speculative execution. We study [27] how it is possible to use memory dependency speculation during the DBT process. Our approach enables fine-grained speculation optimizations thanks to a combination of hardware and software. Our results show that our approach leads to a geo-mean speed-up of 10 % at the price of a 7 % area overhead.

Our previous work on Hybrid-DBT was also presented at the RISC-V workshop in Zürich, Switzerland [38].

This work is a collaboration with the CAIRN team.

7.1.7. Scalable program tracing

Participants: Byron Hawkins, Erven Rohou.

The initial goal of scalable tracing is to record long executions at under 5× overhead (ideally 2×), but it is equally important for analysis of the compressed trace to be efficient. This requires careful organization of the recorded data structures so that essential factors can be accessed without decompressing the trace or comprehensively iterating its paths. Precise context sensitivity is especially important for both optimization and security applications of trace-based program analysis, but scalability becomes challenging for frequently invoked functions that have a high degree of internal complexity. To avoid state space explosion in the context graph, such a function can be represented as a singleton while its complexity is preserved orthogonally. The current efforts focus mainly on developing an integration strategy to simplify program analysis over these two orthogonal dimensions of the trace.

7.1.8. Compiler optimization for quantum architectures

Participant: Caroline Collange.

In 2016, the first quantum processors have been made available to the general public. The possibility of programming an actual quantum device has elicited much enthusiasm [34]. Yet, such possibility also brought challenges. One challenge is the so called Qubit Allocation problem: the mapping of a virtual quantum circuit into an actual quantum architecture. There exist solutions to this problem; however, in our opinion, they fail to capitalize on decades of improvements on graph theory.

In collaboration with the Federal University of Minas Gerais, Brazil, we show how to model qubit allocation as the combination of Subgraph Isomorphism and Token Swapping [31]. This idea has been made possible by the publication of an approximative solution to the latter problem in 2016. We have compared our algorithm against five other qubit allocators, all independently designed in the last two years, including the winner of the IBM Challenge. When evaluated in “Tokyo”, a quantum architecture with 20 qubits, our technique outperforms these state-of-the-art approaches in terms of the quality of the solutions that it finds and the amount of memory that it uses, while showing practical runtime.

7.2. Processor Architecture

Participants: Arthur Blanleuil, Niloofar Charmchi, Caroline Collange, Kleovoulos Kalaitzidis, Pierre Michaud, Anis Peysieux, Daniel Rodrigues Carvalho, André Seznec.

7.2.1. Value prediction

Participants: Kleovoulos Kalaitzidis, André Seznec.

Modern context-based value predictors tightly associate recurring values with instructions and contexts by building confidence upon them [9]. However, when execution monotony exists in the form of intervals, the potential prediction coverage is limited, since prediction confidence is reset at the beginning of each new interval. In [25], we address this challenge by introducing the notion of Equality Prediction (EP), which represents the binary facet of value prediction. Following a two fold decision scheme (similar to branch prediction), EP makes use of control-flow history to determine equality between the last committed result read at fetch time, and the result of the fetched occurrence. When equality is predicted with high confidence, the read value is used. Our experiments show that this technique obtains the same level of performance as previously proposed state-of-the-art context-based predictors. However, by virtue of better exploiting patterns of interval equality, our design complements the established way that value prediction is performed, and when combined with contemporary prediction models, improves the delivered speedup by 19 % on average.

7.2.2. Compressed caches

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

The speed gap between CPU and memory is impairing performance. Cache compression and hardware prefetching are two techniques that could confront this bottleneck by decreasing last level cache misses. However, compression and prefetching have positive interactions, as prefetching benefits from higher cache capacity and compression increases the effective cache size. We propose Compressed cache Layout Aware Prefetching (CLAP) to leverage the recently proposed sector-based compressed cache layouts such as SCC or YACC to create a synergy between compressed cache and prefetching. The idea of this approach is to prefetch contiguous blocks that can be compressed and co-allocated together with the requested block on a miss access [33]. Prefetched blocks that share storage with existing blocks do not need to evict a valid existing entry; therefore, CLAP avoids cache pollution. In order to decide the co-allocatable blocks to prefetch, we propose a compression predictor. Based on our experimental evaluations, CLAP reduces the number of cache misses by 12 % and improves performance by 4 % on average, comparing to a compressed cache [23].

7.2.3. Deep microarchitecture

Participants: Anis Peysieux, André Seznec.

The design of an efficient out-of-order execution core is particularly challenging. When the issue-width increases, the cost of the extra logic required by out-of-core execution increases dramatically. The silicon area occupied by this OoO core tends to grow quasi-quadratically with the issue-width (e.g. issue logic, register file and result bypass). At the same time, the power requirement and the energy consumption of the out-of–order core grow super-linearly with issue width. On wide-issue out-of-order execution cores, issue logic response time, register file access time, as well as result bypass delays represent potential critical paths that might impair cycle time or might necessitate further deepening of the execution pipeline. The objective of the PhD thesis of Anis Peysieux will be to reduce the number of instructions that enter the OoO core, and therefore to master the hardware complexity while still achieving the performance promises of a very wide issue processor.

7.2.4. Dynamic thermal management

Participant: Pierre Michaud.

As power dissipation and circuit temperature constrain their performance, modern processors feature turbo control mechanisms to adjust the voltage and clock frequency dynamically so that circuit temperature stays below a certain limit. In particular, turbo control exploits the fact that, after a long period of low processor activity, the thermal capacity of the chip, its package and the heatsink can absorb heat at a relatively fast rate during a certain time, before the temperature limit constrains that rate. Hence power dissipation can be temporarily boosted above the average sustainable value. The turbo control must monitor circuit temperature continuously to maximize the clock frequency. Temperature can be monitored by reading the integrated thermal sensors. However, making the clock frequency depend on thermal sensor readings implies that processor performance depends on ambient temperature. Yet this form of performance non-determinism is a problem for certain processor makers. A possible solution is to determine the clock frequency not from the true temperature but from a thermal model based on the nominal ambient temperature. Such model should be as accurate as possible in order to prevent sensor-based protection from triggering but sporadically, without hurting performance by overestimating temperature too much. The model should also be simple enough to provide calculated temperature in real time. We propose a thermal model possessing these qualities, and a new turbo control algorithm based on that model [37].

7.2.5. Thread convergence prediction for general-purpose SIMT architectures

Participants: Arthur Blanleuil, Caroline Collange.

GPUs group threads of SPMD programs in warps and synchronize them to execute the same instruction at the same time. This execution model, referred to as Single-Instruction, Multiple-Thread (SIMT), enables the use of energy-efficient SIMD execution units by factoring out control logic such as instruction fetch and decode pipeline stages for a whole warp. SIMT execution is the key enabler for the energy efficiency of GPUs. We seek to generalize the SIMT execution model to general-purpose superscalar cores.

As threads within a warp may follow different directions through conditional branches in the program, the warp must follow each taken path in turn, while disabling individual threads that do not participate. Following divergence, current GPU architectures attempt to restore convergence at the earliest program point following static annotations in the binary. However, this policy has been shown to be suboptimal in many cases, in which later convergence improves performance. In fact, optimal convergence points depend on dynamic program behavior, so static decisions are unable to capture them.

The goal of the thesis of Arthur Blanleuil is to design predictors that enable the microarchitecture to infer dynamic code behavior and place convergence points appropriately. Convergence predictors have analogies with branch predictors and control independence predictors studied in superscalar processor architecture, but they present one additional challenge: the thread runaway problem. Although a branch misprediction will be identified and repaired locally, a wrong thread scheduling decision may go unnoticed and delay convergence by thousands of instructions. To address the thread runaway problem, we plan to explore promise-based speculation and recovery strategies. When no information is available, we follow the traditional conservative earliest-convergence scheduling policy. Once the predictor has enough information to make a more aggressive prediction, it generates assumptions about the prediction. The microarchitecture then keeps checking dynamically whether the assumptions actually hold true in the near future. If assumptions turn out to be wrong, the prediction will be reconsidered by changing back priorities to conservative. Such promise-based speculation policies can address the thread runaway problem by fixing a bound on the worst-case performance degradation of an aggressive scheduling policy against the conservative baseline.

Accurate thread convergence policies will enable dynamic vectorization to adapt to application characteristics dynamically. They will both improve performance and simplify programming of many-core architectures by alleviating the need for advanced code tuning by expert programmers.

7.2.6. Exploring the design space of GPU architectures

Participants: Alexandre Kouyoumdjian, Caroline Collange.

We study tradeoffs in the internal organization of GPUs in the context of general-purpose parallel processing [35]. In particular, we analyze the performance impact of having a few wide streaming multiprocessors compared to many narrow ones. Although we find narrow configurations usually give higher performance for an equal number of execution units, they require more hardware resources and energy. On the other hand, our evaluation show that the optimal streaming multiprocessor width varies across applications. This study motivates adaptive GPU architectures that would support configurable internal organization.

7.3. WCET estimation and optimization

Participants: Loïc Besnard, Damien Hardy, Isabelle Puaut, Stefanos Skalistis.

7.3.1. WCET estimation for many core processors

Participants: Damien Hardy, Isabelle Puaut, 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 [18].

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.

In [28], we propose a scheduling technique that jointly selects Scratchpad Memory (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 technique on streaming applications and synthetic task-graphs. The overlapping of communications with computations allows the length of generated schedules to be reduced by 4 % on average on streaming applications, with a maximum of 16 %, and by 8 % on average 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).

7.3.1.3. Interference-sensitive run-time adaptation of time-triggered schedules

In time-critical systems, run-time adaptation is required to improve the performance of time-triggered execution, derived based on Worst-Case Execution Time (WCET) of tasks. By improving performance, the systems can provide higher Quality-of-Service, in safety-critical systems, or execute other best-effort applications, in mixed-critical systems. To achieve this goal, we propose in [32] a parallel interference-sensitive run-time adaptation mechanism that enables a fine-grained synchronisation among cores. Since the run-time adaptation of offline solutions can potentially violate the timing guarantees, we present the Response-Time Analysis (RTA) of the proposed mechanism showing that the system execution is free of timing-anomalies. The RTA takes into account the timing behavior of the proposed mechanism and its associated WCET. To support our contribution, we evaluate the behavior and the scalability of the proposed approach for different application types and execution configurations on the 8-core Texas Instruments TMS320C6678 platform. The obtained results show significant performance improvement compared to state-of-the-art centralized approaches.

7.3.1.4. 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 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 [19] we present the overall results of the project, a compiler tool-flow for automated parallelization of model-based real-time software, which addresses the shortcomings of multi-core architectures in real-time systems. The flow is demonstrated using a model-based Terrain Awareness and Warning Systems (TAWS) and an edge detection algorithm from the image-processing domain. Model-based applications are first transformed into real-time C code and from there into a well-predictable parallel C program. Tight bounds for the Worst-Case Execution Time (WCET) of the parallelized program can be determined using an integrated multi-core WCET analysis. Thanks to the use of an architecture description language, the general approach is applicable to a wider range of target platforms. An experimental evaluation for a research architecture with network-on-chip (NoC) interconnect shows that the parallel WCET of the TAWS application can be improved by factor 1.77 using the presented compiler tools.

7.3.2. WCET estimation and optimizing compilers

Participants: Isabelle Puaut, Stefanos Skalistis.

Static Worst-Case Execution Time (WCET) estimation techniques operate upon the binary code of a program in order to provide the necessary input for schedulability analysis techniques. Compilers used to generate this binary code include tens of optimizations, that can radically change the flow information of the program. Such information is hard to be maintained across optimization passes and may render automatic extraction of important flow information, such as loop bounds, impossible. Thus, compiler optimizations, especially the sophisticated optimizations of mainstream compilers, are typically avoided. We explore [24] for the first time iterative-compilation techniques that reconcile compiler optimizations and static WCET estimation. We propose a novel learning technique that selects sequences of optimizations that minimize the WCET estimate of a given program. We experimentally evaluate the proposed technique using an industrial WCET estimation tool (AbsInt aiT) over a set of 46 benchmarks from four different benchmarks suites, including reference WCET benchmark applications, image processing kernels and telecommunication applications. Experimental results show that WCET estimates are reduced on average by 20.3 % using the proposed technique, as compared to the best compiler optimization level applicable.

7.3.3. WCET estimation and processor micro-architecture

Participant: Isabelle Puaut.

Cache memories in modern embedded processors are known to improve average memory access performance. Unfortunately, they are also known to represent a major source of unpredictability for hard real-time workload. One of the main limitations of typical caches is that content selection and replacement is entirely performed in hardware. As such, it is hard to control the cache behavior in software to favor caching of blocks that are known to have an impact on an application's worst-case execution time (WCET). In [26], we consider a cache replacement policy, namely DM-LRU, that allows system designers to prioritize caching of memory blocks that are known to have an important impact on an application's WCET. Considering a single-core, single-level cache hierarchy, we describe an abstract interpretation-based timing analysis for DM-LRU. We implement the proposed analysis in a self-contained toolkit and study its qualitative properties on a set of representative benchmarks. Apart from being useful to compute the WCET when DM-LRU or similar policies are used, the proposed analysis can allow designers to perform WCET impact-aware selection of content to be retained in cache.

Long pipelines need good branch predictors to keep the pipeline running. Current branch predictors are optimized for the average case, which might not be a good fit for real-time systems and worst-case execution time analysis. We present [29] a time-predictable branch predictor co-designed with the associated worst-case execution time analysis. Thee branch predictor uses a fully-associative cache to track branch outcomes and destination addresses. The fully-associative cache avoids any false sharing of entries between branches. Therefore, we can analyze program scopes that contain a number of branches lower than or equal to the number of branches in the prediction table. Experimental results show that the worst-case execution time bounds of programs using the proposed predictor are lower than using static branch predictors at a moderate hardware cost.

7.4. Security

Participants: Nicolas Bellec, Damien Hardy, Kévin Le Bon, Isabelle Puaut, Erven Rohou.

7.4.1. Attack detection co-processor for real-time systems

Participants: Nicolas Bellec, Isabelle Puaut.

Real-time embedded systems (RTES) are required to interact more and more with their environment, thereby increasing their attack surface. Recent security breaches on car brakes and other critical components, have already proven the feasibility of attacks on RTES. Such attacks may change the control-flow of the programs, which may lead to violations of the timing constraints of the system. In this ongoing work, we design a technique to detect attacks in RTES based on timing information. Our technique is based on a monitor, implemented in hardware to preserve the predictability of instrumented programs. The monitor uses timing information (Worst-Case Execution Time – WCET – of code regions) to detect attacks. An algorithm for the region selection, optimal when the monitoring memory is not limited is presented and provides guarantees on attack detection latency. An implementation of the hardware monitor and its simulation demonstrates the practicality of our approach. An experimental study evaluates the maximum attack detection latency for different monitor memory budgets.

This work is done in collaboration with the CIDRE and CAIRN teams.

7.4.2. Multi-nop fault injection attack

Participants: Damien Hardy, Erven Rohou.

The CIDRE team has developed a platform named Traitor that allows to perform multiple fault injection attack by replacing instructions by nops during the execution of a program. In this context, we are defining a program model where each instruction can be replaced by a nop at runtime. On this model we plan to apply compilation techniques on the binary to automatically determine where nops have to be inserted at runtime to perform sophisticated attacks such as dump of memory, modification of the memory, memory protection deactivation, execution of code in RAM.

This work is done in collaboration with the CIDRE team.

7.4.3. Compiler-based automation of side-channel countermeasures

Participants: Damien Hardy, Erven Rohou.

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 proposed a compiler extension to automatically apply different masking schemes for block cipher algorithms. We introduced a generic approach to describe the schemes and we inserted three of them at compile-time on an AES implementation. Currently, a practical side-channel analysis is performed in collaboration with TAMIS to assess the correctness and the performance of the code inserted.

This work is done in collaboration with the TAMIS team.

7.4.4. Platform for adaptive dynamic protection of programs

Participants: Kévin Le Bon, Erven Rohou.

Memory corruption attacks are a serious threat for system integrity. Many techniques have been developed in order to protect systems from these attacks. However, the deployment of heavy protections often degrades the performance of programs. We propose [36] a dynamic approach that adapts the protection level of the target process during its execution depending on the observed behavior.

HYCOMES Project-Team

6. New Results

6.1. Mathematical Foundations of Physical Systems Modeling Languages

Participants: Albert Benveniste, Benoît Caillaud, Mathias Malandain.

Modern modeling languages for general physical systems, such as Modelica or Simscape, rely on Differential Algebraic Equations (DAE), i.e., constraints of the form $f(\dot{x}, x, u) = 0$. This facilitates modeling from first principles of the physics. This year we completed the development of the mathematical theory needed to sound, on solid mathematical bases, the design of compilers and tools for DAE based physical modeling languages.

Unlike Ordinary Differential Equations (ODE, of the form $\dot{x} = g(x, u)$), DAE exhibit subtle issues because of the notion of *differentiation index* and related *latent equations*—ODE are DAE of index zero for which no latent equation needs to be considered. Prior to generating execution code and calling solvers, the compilation of such languages requires a nontrivial *structural analysis* step that reduces the differentiation index to a level acceptable by DAE solvers.

Multimode DAE systems, having multiple modes with mode-dependent dynamics and state-dependent mode switching, are much harder to deal with. The main difficulty is the handling of the events of mode change. Unfortunately, the large literature devoted to the numerical analysis of DAEs does not cover the multimode case, typically saying nothing about mode changes. This lack of foundations causes numerous difficulties to the existing modeling tools. Some models are well handled, others are not, with no clear boundary between the two classes. Basically, no tool exists that performs a correct structural analysis taking multiple modes and mode changes into account.

In our work, we developed a comprehensive mathematical approach supporting compilation and code generation for this class of languages. Its core is the *structural analysis of multimode DAE systems,* taking both multiple modes and mode changes into account. As a byproduct of this structural analysis, we propose well sound criteria for accepting or rejecting models at compile time.

For our mathematical development, we rely on *nonstandard analysis,* which allows us to cast hybrid systems dynamics to discrete time dynamics with infinitesimal step size, thus providing a uniform framework for handling both continuous dynamics and mode change events.

A big comprehensive document has been written, which will be finalized and submitted next year.

6.2. Structural analysis of multimode DAE systems

Participants: Albert Benveniste, Benoît Caillaud, Khalil Ghorbal, Mathias Malandain.

The Hycomes team has obtained two results related to the structural analysis of multimode DAE systems.

6.2.1. Impulsive behavior of multimode DAE systems

A major difficulty with multimode DAE systems are the commutations from one mode to another one when the number of equations may change and variables may exhibit impulsive behavior, meaning that not only the trajectory of the system may be discontinuous, but moreover, some variables may be Dirac measures at the instant of mode changes. In [7] , we compare two radically different approaches to the structural analysis problem of mode changes. The first one is a classical approach, for a restricted class of DAE systems, for which the existence and uniqueness of an impulsive state jump is proved. The second approach is based on nonstandard analysis and is proved to generalize the former approach, to a larger class of multimode DAE systems. The most interesting feature of the latter approach is that it defines the state-jump as the standardization of the solution of a system of system of difference equations, in the framework of nonstandard analysis.

6.2.2. An implicit structural analysis method for multimode DAE systems

Modeling languages and tools based on Differential Algebraic Equations (DAE) bring several specific issues that do not exist with modeling languages based on Ordinary Differential Equations. The main problem is the determination of the differentiation index and latent equations. Prior to generating simulation code and calling solvers, the compilation of a model requires a structural analysis step, which reduces the differentiation index to a level acceptable by numerical solvers.

The Modelica language, among others, allows hybrid models with multiple modes, mode-dependent dynamics and state-dependent mode switching. These Multimode DAE (mDAE) systems are much harder to deal with. The main difficulties are (i) the combinatorial explosion of the number of modes, and (ii) the correct handling of mode switchings.

The focus of the paper [31] is on the first issue, namely: How can one perform a structural analysis of an mDAE in all possible modes, without enumerating these modes? A structural analysis algorithm for mDAE systems is presented, based on an implicit representation of the varying structure of an mDAE. It generalizes J. Pryce's Σ-method [56] to the multimode case and uses Binary Decision Diagrams (BDD) to represent the mode-dependent structure of an mDAE. The algorithm determines, as a function of the mode, the set of latent equations, the leading variables and the state vector. This is then used to compute a mode-dependent block-triangular decomposition of the system, that can be used to generate simulation code with a mode-dependent scheduling of the blocks of equations.

This method has been implemented in the IsamDAE software. This has allowed the Hycomes team to evaluate the performance and scalability of the method on several examples. In particular, it has been possible to perform the structural analysis of systems with more than 750 equations and 10^{23} modes.

6.3. Functional Decision Diagrams: A Unifying Data Structure For Binary Decision Diagrams

Participants: Joan Thibault, Khalil Ghorbal.

Zero-suppressed binary Decision Diagram (ZDD) is a notable alternative data structure of Reduced Ordered Binary Decision Diagram (ROBDD) that achieves a better size compression rate for Boolean functions that evaluate to zero almost everywhere. Deciding *a priori* which variant is more suitable to represent a given Boolean function is as hard as constructing the diagrams themselves. Moreover, converting a ZDD to a ROBDD (or vice versa) often has a prohibitive cost. This observation could be in fact stated about almost all existing BDD variants as it essentially stems from the non-compatibility of the reduction rules used to build such diagrams. Indeed, they are neither interchangeable nor composable. In [8], we investigate a novel functional framework, termed Lambda Decision Diagram (LDD), that ambitions to classify the already existing variants as implementations of special LDD models while suggesting, in a principled way, new models that exploit application-dependant properties to further reduce the diagram's size. We show how the reduction rules we use locally capture the global impact of each variable on the output of the entire function. Such knowledge suggests a variable ordering that sharply contrasts with the static fixed global ordering in the already existing variants as well as the dynamic reordering techniques commonly used.

Kairos Project-Team

7. New Results

7.1. Spatio-temporal constraints for mobile systems, with automotive driving assistance illustrations

Participants: Frédéric Mallet, Joëlle Abou Faysal, Robert de Simone, Xiaohong Chen.

The objective here is to extend constraint specifications to encompass spatial aspects in addition to logical multiform time. Spatio-temporal logics and requirement formalisms are thus an inspiration here. But mobility requests additionally that these spatio-temporal relations evolve in time. We are investigating in several directions:

- a target methodological approach is to consider these spatio-temporal relations to express safe driving rules as requirements or guarantees, meant to (in)validate trajectory proposals computed by a lower-level algorithmic system (itself operating on more direct neighborhood information). A realistic size case study is handled in collaboration with Renault Software Labs, as part of the CIFRE PhD contract of Joëlle Abou-Faysal, to define the precise needs in expressiveness and formal validation.
- Preliminary definitions of a spatio-temporal requirement specification languages, borrowing ideas from spatio-temporal logics and formal mobile process modeling (none of which being sufficient to our aim), is being progressed in collaboration with fellow researchers from ECNU Shanghai [20].

7.2. System Engineering for Performance and Availability in satellite embedded COTS

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

In the context of the IRT ATIPPIC project, which provided engineer position funding for Paul Bouche and Amin Oueslati, we investigated the application of a realistic formal design methodology applied on a real case study under construction by the ATIPPIC partners, in this case a prototype satellite based on general-purpose electronic Components-on-the-Shelf (COTS), not radiation-hardened. We focused on the one hand on the Model-Based Design of local interconnects, to provide analysis techniques regarding bandwidth and possible congestion of inter-process communications; on the other hand, we considered formal analysis of availability in case of fault (solar radiations), to study impact of alternative mitigation techniques for fault tolerance. Results were delivered in the form of Capella viewpoints and analysis tools to the IRT Saint-Exupéry, as free software. They were also published in [18], [23].

7.3. Efficient solvers and provers for CCSL

Participants: Frédéric Mallet, Xiaohong Chen.

One of the goal of the team is to promote the use of logical time in various application domains. This requires to have efficient solvers for CCSL. We have made considerable progresses on this part along two lines. One by relying on SMT solvers (like Z3), the other by building a dynamic logic amenable to building semi-automatic proofs for logical time properties of reactive systems. Then for some classes of problems we can efficient solving tools.

- The first step is to have an efficient Z3 library for solving CCSL specifications. We have improved a lot the performances over last year by getting rid of some of the existential quantifiers in our properties [35].

- Second, we use this solver to help requirement engineers elicite the requirements. We use execution traces to help generate valid satisfied CCSL specifications [28].
- Third, we have built a dynamic logics based on CCSL, where the formulae are derived from CCSL relational operators and programs include some of CCSL expressions and some imperative reactive constructs akin to Esterel programs. Then we have an interactive proof system, that helps prove that some reactive program satisfies a set of formulas at all time. As we use only a subset of CCSL then, we can restrict to a decidable subset of the logics and the SMT solver is always efficient. The SMT helps guide the semi-automatic proof [34] by identifying the next proof rules that can be used (or not).

7.4. Formal temporal Smart Contracts

Participants: Frédéric Mallet, Marie-Agnès Peraldi Frati, Robert de Simone.

"Smart Contracts", as a way to define legal ledger evolution in Blockchain environments, can be seen as rule constraints to be satisfied by the set of their participants. Such contracts are often reflecting requirements or guarantees extracted from a legal or financial corpus of rules, while this can be carried to other technical fields of expertise. Our view is that Smart Contracts are often relying on logically timed events, thus welcoming the specification style of our formalisms (such as CCSL). The specialization of multiform logical time constraints to this domain is under study, in collaboration with local academic partners at UCA UMR LEAT and Gredeg, and industrial partners, such as Symag and Renault Software Labs. Local funding was obtained from UCA DS4H EUR Academy 1, which allowed preparation of the ANR project SIM that was accepted in 2019. One goal is to get acceptance from the lawyers while still preserving strong semantics for verification. This builds on our previous expertise [16].

7.5. CCSL extension to Stochastic logical time

Participants: Frédéric Mallet, Robert de Simone.

CCSL specifications allows distinct clocks with unfixed inter-relations. In settings such as cyber-physical modeling, probabilistic rates of relative occurrences may be provided as bounds. The objective is to provide construct to introduce such relations for the inclusion and precedence partial orders, but also to consider also constructs that associate them. Preliminary results have been obtained by Frédéric Mallet in collaboration with fellow researchers from ECNU Shanghai.

7.6. Semantic Resource Discovery in Internet

Participant: Luigi Liquori.

Results [30] are obtained in close collaboration with professors Matteo Sereno and Rossano Gaeta from the University of Turin. Internet in recent years has become a huge set of channels for content distribution highlighting limits and inefficiencies of the current protocol suite originally designed for host-to-host communication. We propose a Content Name System Service (CNS) that provides a new network aware Content Discovery Service. The CNS behavior and architecture uses the BGP inter-domain routing information. In particular, the service registers and discovers resource names in each Internet Autonomous System (AS): contents are discovered by searching through the augmented AS graph representation classifying ASes into customer, provider, and peering, as the BGP protocol does. An interesting extension of this Internet Service could be to scale up to Internet of Things and to Cyber Physical Systems inter-networked with networks different than Internet.

7.7. Raising Semantic Resource Discovery in IoT

Participants: Luigi Liquori, Marie-Agnès Peraldi Frati.

Within the standards for M2M and the Internet of Things, managed by ETSI, oneM2M, we are looking for suitable mechanisms and protocols to perform a Semantic Resource Discovery as described in the previous Subsection. More precisely, we are extending the (actually weak) Semantic Discovery mechanism of the IoT oneM2M standard. The goal is to enable an easy and efficient discovery of information and a proper internetworking with external source/consumers of information (e.g. a data bases in a smart city or in a firm), or to directly search information in the oneM2M system for big data purposes. oneM2M ETSI standard has currently a rather weak native discovery capabilities that work properly only if the search is related to specific known sources of information (e.g. searching for the values of a known set of containers) or if the discovery is very well scoped and designed (e.g. the lights in a house). We submitted our vision in ETSI project submission “Semantic Discovery and Query in oneM2M” (currently under evaluation by ETSI) for extending oneM2M with a powerful Semantic Resource Discovery Service, taking into account additional constraints, such as topology, mobility (in space), intermittence (in time), scoping, routing ...

7.8. Empirical study of Amdahl’s law on multicore processors

Participants: Carsten Bruns, Sid Touati.

Since many years, we observe a shift from classical multiprocessor systems to multicores, which tightly integrate multiple CPU cores on a single die or package. This shift does not modify the fundamentals of parallel programming, but makes harder the understanding and the tuning of the performances of parallel applications. Multicores technology leads to sharing of microarchitectural resources between the individual cores, which Abel et al. classified in storage and bandwidth resources. In this research report [39], we empirically analyze the effects of such sharing on program performance, through repeatable experiments. We show that they can dominate scaling behavior, besides the effects described by Amdahl’s law and synchronization or communication considerations. In addition to the classification of Abel et al., we view the physical temperature and power budget also as a shared resource. It is a very important factor for performance nowadays, since DVFS over a wide range is needed to meet these constraints in multicores. Furthermore, we demonstrate that resource sharing not just leads a flat speedup curve with increasing thread count but can even cause slowdowns. Last, we propose a formal modeling of the performances to allow deeper analysis. Our work aims to gain a better understanding of performance limiting factors in high performance multicores, it shall serve as basis to avoid them and to find solutions to tune the parallel applications.

7.9. Communicating Networks of Data-Flow (sub)networks with limited memory

Participant: Robert de Simone.

Process Networks have been proposed a long time ago as models of concurrent, embedded streaming computations and communications, both amenable to formal analysis as models and executable as parallel program abstractions. As part of a larger effort at identifying precise connections between these models, programming models, and embedded parallel architectures altogether, we worked this year on the following problem: given a network of concurrent processes (Kahn-style) where each process is in turn a data-flow process network (SDF-style), can we decide in an efficient fashion (not NP-hard) whether a given assignment of communications to bounded local memories is schedulable (in a way that two simultaneous communications cannot require more than the available memory). A technical report is in preparation.

7.10. Behavioral Equivalence of Open Systems

Participants: Eric Madelaine, Cristian Grigoriu, Zechen Hou.

We consider Open (concurrent) Systems where the environment is represented as a number of processes which behavior is unspecified. Defining their behavioral semantics and equivalences from a Model-Based Design perspective 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. In collaboration with ENS Lyon and Inria Lille, we studied an application of this approach to the verification of BIP architectures; this work extends previous dedicated approaches for compositional verification of BIP systems to data-dependent synchronizations [22]. Following last year results we have devised dedicated algorithms for checking equivalence of such systems [27], [41], currently under implementation in collaboration with ECNU Shanghai.

In order to facilitate the usage of our tools, we have also defined a language for defining open systems in terms of parameterized networks of synchronized automata (pNets, [4]), and implemented this language as an Eclipse-based editor in the VerCors tool (see Software section), together with interfaces to the semantic construction and equivalence checking algorithms.

7.11. Calculi with Union and Intersection types

Participants: Luigi Liquori, Claude Stolze.

Union and intersection types are interesting to improve actual programming languages static disciplines with alternative form of polymorphism. Since type inference is undecidable, our research vein focus on finding suitable "type decorations" in term syntax permitting to make type checking decidable, *i.e.*$\lambda x.x : (\sigma \to \sigma) \cap (\tau \to \tau)$ becomes $\langle \lambda x : \sigma.x, \lambda x : \tau.x \rangle : (\sigma \to \sigma) \cap (\tau \to \tau)$ in a fully-typed syntax. Those type systems uses intensively a subtyping relation stating *e.g.* that $\sigma \cap \tau \leq \sigma$ or $\sigma \leq \sigma \cup \tau$. Deciding whether $\sigma \leq \tau$ can be extremely difficult in complexity (space and time): actually, there are few algorithms in the literature dealing with union and intersection types. Recently [45] we have proved and certified in Coq a subtype algorithm of a type theory with union and intersection types; we have also extracted a running functional code. Subtyping constraints could be easily interpreted as temporal constraints in a suitable temporal algebra, like those that could be specified in CCSL. Advances of typed-calculi featuring those type disciplines are presented in [42], [31] and [14].

7.12. Bull, an Interactive Type Checker with Union and Intersection Types

Participants: Luigi Liquori, Claude Stolze.

Starting from our theoretical researches on Intersection and Union Types and related Subtype Theories, we have designed and implemented a prototype of an Interactive Typechecker based on the 2018 work on the Δ-framework [43], on the 2017 work on decidable subtyping logic for Intersection and Union types [45], and on our recent advances on the Δ-calculus [42] and [14]. The prototype is called *Bull*; Bull has a command-line interface where the user can declare axioms, terms, and perform computations. These terms can be incomplete, therefore the type checking algorithm uses unification to try to construct the missing subterms. A Read-Eval-Print-Loop allows to define axioms and definitions, and performs some basic terminal-style features like error pretty-printing, subexpressions highlighting, and file loading. Moreover, it can typecheck a proof and normalize it. We use the syntax of *Pure Type Systems* of Berardi to improve the compactness and the modularity of the kernel. Abstract and concrete syntax are mostly aligned: the concrete syntax is similar to the concrete syntax of the ITP Coq. We have also designed and implemented a *higher-order unification algorithmà la* Huet for terms, while typechecking and partial type inference are done by our *bidirectional refinement algorithm*. The refinement can be split into two parts: the essence refinement and the typing refinement. The bidirectional refinement algorithm aims to have partial type inference, and to give as much information as possible to the unifier. For instance, if we want to find a $?y$ such that $\vdash_\Sigma \langle \lambda x : \sigma.x, \lambda x : \tau.?y \rangle : (\sigma \to \sigma) \cap (\tau \to \tau)$, we can infer that $x : \tau \vdash ?y : \tau$ and that $\wr ?y \wr =_\beta x$. We are experimenting with classical examples in Bull, like the ones formalized by Pfenning with his Refinement Types in LF, and we are looking for examples taking into account preorders, constraints and operators (like *e.g.*$<, \leq, >, \geq, \cup, \cap$...) that could be interpreted as timed algebras expressions *à la* CCSL. This would be a little step toward the formal and certified definition of a simple timed type systems for the λ-calculus and a Timed Logical Framework.

The software can be actually retrieved on the GitHub repository Bull (registration to the BIL Inria data base is in progress).

7.13. Co-Modeling for Better Co-Simulations

Participants: Julien Deantoni, Giovanni Liboni.

A Collaborative simulation consists in coordinating the execution of heterogeneous models executed by different tools. In most of the approaches from the state of the art, the coordination is unaware of the behavioral semantics of the different models under execution; *i.e.*, each model and the tool to execute it is seen as a black box. We highlighted that it introduces performance and accuracy problems [44].

In order to improve the performance and correctness of co-simulations, we proposed a language to defined model behavioral interfaces, *i.e.*, to expose some information about the model behavioral semantics. We also proposed another language to make explicit the way to coordinate the different models by using dedicated connectors. The goal is to provide few information about the models to avoid intellectual property violations, but enough to allow an expert to make relevant choices concerning their coordination. The resulting models can then be exploited to generate a dedicated coordination, aware of the specificity of each model [29]. Future work mainly consists in experimenting a new co-simulation interface taking advantage of the model behavioral interface and proposed as a generalization of co-simulation interfaces from the state of the art.

This work is realized in the context of the GLOSE project (see Section 1) in collaboration with Safran and other Inria teams (namely HyCOMES and DiVerSE).

7.14. CCSL for Models Behavioral Composition

Participants: Julien Deantoni, Frédéric Mallet, Hui Zhao.

The growing use of models for separating concerns in complex systems has lead to a proliferation of model composition operators. These composition operators have traditionally been defined from scratch following various approaches differing in formality, level of detail, chosen paradigm, and styles. Due to the lack of proper foundations for defining model composition (concepts, abstractions, or frameworks), it is difficult to compare or reuse composition operators. In [17], we proposed research directions towards a unifying framework that reduces all structural composition operators to structural merging, and all composition operators acting on discrete behaviors to event scheduling. Our belief is that CCSL, embedding both synchronous and asynchronous relations, is a good candidate to specify the event scheduling corresponding to the coordination of the different behaviors. However, as already stated in previous sections, to achieve such a status, some extensions to CCSL must be proposed. One of them was the possibility to prioritize events in the presence of synchronous relations. This was formally defined in [26] and implemented in the TimeSquare tool.Other interesting extensions are under study in the context of heterogeneous models, see Section 7.13 .

As part of Zhao Hui's PhD work, we have proposed a language to bring together subsets of existing predefined languages in a bid to combine their expressiveness. Rather than trying to build the ultimate unified language, sum of all languages, we would rather select meaningful features in existing languages and build a new language based on those features. As an example of application, we have shown how to combine the functional models of Capella with the security models of SysML-sec in an ad-hoc security-aware language for functional analysis [36].

7.15. Expressing IoT security constraints

Participants: Stéphanie Challita, Robert de Simone.

In the framework of Inria Project Lab SPAI, we are considering extensions of the logical time constraint style of CCSL, in order to encompass locality information as well as the duality between (dynamic) agents and (static) resources. Once an appropriate framework has been defined to express occupancy of resources by agents through (logical) time, notions of access rights, enclaves, privileges and priorities may be encoded straightforwardly, and rules governing their proper secure use can be expressed as properties. Results will be presented at the completion of Stephanie Challita postdoctoral period.

7.16. Real-Time Systems Compilation

Participants: Dumitru Potop Butucaru, Hugo Pompougnac, Jad Khatib.

This work took place in the framework of the PIA ES3CAP project (see section 9.2.5) and in close collaboration with Inria PARKAS, Airbus, Safran Aircraft Engines, Kalray, and the IRT Saint-Exupéry. It funded the last year of Keryan Didier PhD thesis (before the Paris Kairos subteam was created).

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 one 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) multi-periodic task graphs of real-world control applications, and the problem of statically managing all their computational, memory, synchronization and communication resources.

Our work has provided 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 a 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. Orchestration is not a mere combination of existing tools and algorithms. Enabling predictable execution and keeping pessimism under control requires the formal and algorithmic integration of all design phases, which in turn required the definition of an application normalization phase that facilitates timing analysis, of an original code generation algorithm designed to provide mapping-independent worst-case execution time bounds, and of new real-time scheduling algorithms capable of orchestrating memory allocation and scheduling.

Extensive results on the application of this method to real-file avionics case studies (>5000 unique nodes) mapped on the Kalray MPPA256 Bostan many-core have been presented in [15], [21] and in the PhD thesis of Keryan Didier, defended in September.

The Kalray MPPA platform provides excellent support for safety-critical real-time implementation, by allowing the computation of static WCET bounds. This is no longer true on more classical multi-cores such as those with ARM and POWER micro-architecture. We are currently aiming at extending our method to allow mapping on such multi-cores. Full schedulability guarantees cannot be provided on such platforms. Instead, our aim is to allow the synthesis of implementations that are functionally correct, efficient, and where impredictability is reduced to a minimum by eliminating controllable sources of timing variability. This line of work has been pursued in the context of the collaboration contracts with Airbus and IRT Saint-Exupéry. First results are promising.

Further extensions of our method are under way, most notably to cover timing predictable architectures different from the Kalray MPPA 256.

7.17. Formal Modeling of Concurrent Implementations

Participant: Dumitru Potop Butucaru.

Concurrent programming is notoriously difficult, especially in constrained embedded contexts. Threads, in particular, are wildly non-deterministic as a model of computation, and difficult to analyze in the general case. Fortunately, it is often the case that multi-threaded, semaphore-synchronized embedded software implements high-level functional specifications written in a deterministic data-flow language such as Scade or (safe subsets of) Simulink. We claim that in this case the implementation process should build not just the multi-threaded C code, but (first and foremost) a richer model exposing the data-flow organization of the computations performed by the implementation. From this model, the C code is extracted through selective pretty-printing, while knowledge of the data-flow organization facilitates analysis.

This year, we have proposes a language for describing such implementation models that expose the data-flow behavior hiding under the form of a multi-threaded program. The language allows the representation of efficient implementations featuring pipelined scheduling and optimized memory allocation and synchronization. We showed applicability on a large-scale industrial avionics case study and on a commercial many-core [24].

7.18. Scalability of Constraint Programming for Real-Time Scheduling

Participants: Dumitru Potop Butucaru, Robert de Simone.

Given two abstract modeling descriptions, one of a dataflow process network for the application, one of a block diagram structure for the computing platform and its interconnects, together with cost functions for the elementary computations and communications, one is bound to seek optimal mappings pairing the two. Amongst all the possible techniques, an obvious one consists in using general constraint solvers (real, integer, or boolean constraint programming, SMT solvers, CP solvers, etc.). Given the NP-hard nature of the problem, the issue here is to experimentally determine the empyrical complexity of various scheduling problems, and thus help in determining when solvers can be used for the resolution of scheduling problems.

In previous years we addressed this issue for ILP and SMT solvers. This year, we considered a Constraint Programming solver with dedicated support for modeling and solving real-time scheduling problems (IBM ILOG CPLEX CP Optimizer). The work was conducted in the framework of Bimael Iosif's student internship, and the writing of a paper is under way.

KOPERNIC Team

7. New Results

7.1. Uniprocessor Mixed-Criticality Real-Time Scheduling

In the context of the FUI CEOS project 8.1.1.1 , last two years we transformed the free software program PX4, which performs the autopilot of the CEOS drone, in a graph of hard real-time tasks. This transformation was intended to achieve a schedulability analysis guaranteeing the autopilot is able to perform safety critical missions since its behaviour is deemed to be hard real-time, i.e., all deadlines of all tasks are satisfied. It is worth noting that the autopilot is one of the most important programs of the drone since it maintains its stability not only during hover phases but also during automatic flight missions from one GPS point to another. This transformation resulted in a "real-time autopilot" that we called PX4-RT.

For the first version of PX4-RT we chose, as periods, the periods used in the original version of PX4 which was not hard real-time as we shown last year. Then, since these periods was inherited from an automatic control analysis achieved by initial designers of PX4 in a non hard real-time context, we had to determine the right combination of periods of tasks, allowing on the one hand to correctly control the drone, and on the other hand, using a schedulabilty analysis, to satify all the deadlines. In order to achieve this goal, we used a hardware in the loop simulation (HitL) which simulates only the sensors and the actuators, whereas the PX4-RT program runs on the Pixhawk board based on an ARM Cortex-M4 uniprocessor. Eventually, we determined some period combinations that fit our needs, other combinations did not allow the drone to follow correctly the given mission or resulted in a crash. Moreover, we verified that all the right combinations led to a schedulable set of tasks, meaning that corresponding versions of PX4-RT were hard real-time. Finally, we used with success the best combination of periods to run PX4-RT on the real drone of CEOS during a simple flight. Of course, we plan to achieve numerous realistic flights planned in the three industrial use cases of the CEOS project.

In addition to this study intended to determine the right combination of periods, we addressed two other issues. In the first one we tried to decrease the worst case execution times (WCET) of tasks in order to increase the schedulability ratio. Such decrease allows to add on the same processor new tasks presently executed on other processors, e.g., mission planning, fault tolerance, etc. Since we found out that the Kalman filter had the largest measured execution time of all the tasks, we studied the Kalman filter algorithm implemented in PX4 to decrease its WCET. We suppressed the two states of the Kalman Filter corresponding to the wind speed estimation since our drone do not have a sensor measuring this speed. Then, we suppressed the three states of the Kalman filter corresponding to the accelerometer bias whose standard deviation was close to zero. Each of these modifications brought an improvement of 15 percent in term of largest measured execution time without decreasing the performances of the drone. In the second issue we started a theoretical study about relations between the stability of a set of automatic control laws and the schedulability of the corresponding set of real-time tasks. In the literature some results exist about one control law corresponding to one real-time task. To the best of our knowledge there is no result for a set of control laws that exchange data.

Finally, we deeply studied NuttX the real-time operating system used presently to support PX4 and PX4-RT autopilot programs. Indeed, we plan to modify the scheduler of this operating system in order to manage real-time tasks more safely. In order to do that we will draw inspiration from the technique proposed in our time triggered offline scheduler that accounts for the preemption and scheduler cost [14].

7.2. Multicore processor graph tasks scheduling

Due to widespread of multicore processors on embedded and real-time systems, we concentrate our work on the study of the schedulability of real-time tasks with precedence constraints on such processors. We consider preemptive fixed-priority scheduling policies. First, we have proposed a response time analysis for directed acyclic graphs task model with non-probabilistic execution time and preemptive fixed-priority scheduling policy [10]. Our response time analysis improves importantly the state of the art analyses, while allowing scalable extensions for response time analysis of tasks with worst case execution times described by probability distributions. We extend this response time analysis to similar task model with probabilistic worst case execution time with the advantage of providing efficient results also for task model with non-probabilistic worst case execution times. Our response time analysis is based on iterative equations which offer run-time enhancement compared to existing work [21] requesting the resolution of complex MILP optimization problem. In addition, we have defined priority on sub-task level enhancing the schedulability and reducing the worst-case response time. The proposed priority assignment algorithm is adapted for the studied task model and it outperforms several state-of-the art methods. We have also proposed a partitioning heuristic that assigns each sub-task to a given core. This heuristic takes into consideration communication delays between sub-tasks inside the same graph in order to minimize the communication while balancing different cores load and maximizing possible parallelism. The proposed heuristics and response time analysis (RTA) are validated on randomly generated task sets and on the PX4-RT drone autopilot programs developed by Kopernic team in FR FUI21 CEOS project.

7.3. Power consumption of probabilistic real-time systems

Energy consumption on real-time systems is a crucial problem nowadays as these systems are becoming complex and are expected to deliver more and more functionalities. At the same time, while the processing demand increases, the vast majority of these systems are powered by batteries and are deployed in hazardous environments making their maintenance difficult and impractical. Existing works on energy consumption and real-time systems are often based on a technique called Dynamic Voltage and Frequency Scaling (DVFS). The principle of this technique is to reduce the frequency of the processor in order to lower its input voltage, consequently reducing the energy required to power the processor. Nevertheless, by reducing the frequency of the processor, programs tend to take more time to complete their execution. In the context of real-time systems, programs need to finish their execution before a given deadline. Therefore, the goal of DVFS techniques is to derive proper frequencies that minimize energy consumption and still ensure that all deadlines of all the programs will be respected. Works carried during this postdoc are twofold. The first contribution consisted in observing how the Worst-Case Execution Time (WCET) of programs varies with regards to the frequency of the processor. Many existing works have considered that the WCET is completely scalable, i.e., a simple factor can be applied to derive a new WCET under a different frequency setup. Nevertheless, researchers have recognize that this hypothesis may be too optimistic since other components, that do not run at the same speed as the processor, e.g., the memory, are used by programs. We derived an experimental setup to observe how the execution of programs varied by setting different frequencies on the processor and the memory. We measured CPU cycles and execution times and it was clear from our experiments that the theoretical speedup bound that should be achieved when the processor is running at its maximum speed is never achieved. We also observed, that DVFS techniques could also be applied to the memory of the system, since some programs do not perform many memory request. Our experiments led to a short paper accepted for the Work-in-Progress session of the 40th Real-Time System Symposium. The paper also introduced the task model that will be used as a basis of the next contribution of the postdoc. This next contribution consists in developing RTA techniques for probabilistic real-time systems in order to derive hardware frequency setups. The inclusion of probabilistic real-time system is motivated by the ever-increasing demand of functionalities for this type of systems. To the best of our knowledge, DVFS techniques in conjunction with probabilistic real-time systems have never been studied. The solution to this optimization problem is ongoing work while preparing the submission of first results beginning of February 2020.

7.4. Data-oriented scheduling approaches

We consider the scheduling problem of tasks using an inter-task communication model based on a circular buffer, which eases the data consistency between tasks [13], [12]. The tasks are scheduled on one processor by a fixed priority preemptive scheduling algorithm and they have implicit deadlines. We provide a formal method calculating the optimal size for each of the buffers while ensuring data consistency, i.e., it is required that a buffer slot is accessed for reading the input data. This later slot will never be used by the producer task to write new data before the execution completion of the instances of all consumers that are currently reading from this slot. As a second contribution, we provide an analytical characterization of the temporal validity and reachability properties of the data flowing in between communicating tasks. These two properties are characterized by considering both tasks execution and data propagation orders. Moreover, we assume that a task instance reads all its inputs data at its activation time and writes back the output data at the completion time where this data becomes immediately available for consumption. Given that, they may be several data samples available in the buffer, we say that a data sample is fresh or temporal valid if, since the time instant it is produced, its producer has not completed another execution. Given that, we use buffers whose size may be larger than one, it is obvious that the consumer task will not implicitly know which data is temporally valid. In order to use the data that reflects the current status of the system environment (valid data), we introduce a novel parameter; the sub-sampling rate used within two scheduling algorithms. These scheduling algorithms ensure the data consistency and temporal validity, while deadlines are met.

PARKAS Project-Team

6. New Results

6.1. Efficiently Subtyping Union Types

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. In 2017 with Vitek we started a research project to study the design and the pragmatic use of the Julia language, and formalised the Julia subtyping algorithm. In 2018 we have pursued this study, and we have proved correct the clever and space efficient algorithm relied upon by the Julia runtime. This has been published in [17].

6.2. Fast and reliable unwinding via DWARF tables

Participants: Theophile Bastian, Rémy Oudin, 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.

We have built 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 12x and 25x (depending on application), with a 2.5x size overhead for unwind information.

This work has been published in [13].

6.3. Verified compilation of Lustre

Participants: Timothy Bourke, Lélio Brun, Paul Jeanmaire, Marc Pouzet.

Vélus is a compiler for a subset of LUSTRE and SCADE that is specified in the Coq [28] Interactive Theorem Prover (ITP). It integrates the CompCert C compiler [34], [29] to define the semantics of machine operations (integer addition, floating-point multiplication, etcetera) and to generate assembly code for different architectures. The research challenges are to

- to mechanize, i.e., put into Coq, the semantics of the programming constructs used in modern languages for MBD;
- to implement compilation passes and prove them correct;
- to interactively verify source programs and guarantee that the obtained invariants also hold of the generated code.

This year we created a website for the project (https://velus.inria.fr) and made an initial release under an Inria non-commerical license (https://github.com/Inria/velus). T. Bourke's JCJC ("Jeune Chercheuse Jeune Chercheur") project *FidelR* was accepted for funding by the ANR: it aims to develop ITP-based techniques for treating state machines and interactive program verification. We also made progress on the compilation of the modular reset construct, the treatment of (non-normalized) Lustre, and our longer term goal of strengthening the main correctness theorem. These results are detailed below.

6.3.1. Compiling the modular reset construct.

In the original LUSTRE language, the only way to reset the internal state of an instantiated function is to propagate and test explicit reset signals. Later languages, like LUCID SYNCHRONE and SCADE, provide a construct for resetting an instance modularly (it works for any function) and efficiently (testing occurs only at the point of instantiation). Last year we showed how to encode the semantics of this construct in Coq. This year we focused on its compilation and the associated proof of correctness. We designed and implemented a new intermediate language that exposes different *step* and *reset* actions on node instances. This language facilitates the optimization of conditional statements in the generated code and permits the transformation to imperative code and its proof of correctness to be treated in two steps: one to introduce named memories and another to fix the sequential order of execution. This work forms the core of L. Brun's thesis, to be defended early next year, and an article accepted at the ACM SIGBED international conference on Principles of Programming Languages (POPL 2020).

6.3.2. Non-normalized Lustre.

Our previous work has focused on a subset of "normalized" programs where the form of expressions and equations is constrained to facilitate the compilation. We have generalized the definitions of syntax and semantics in our prototype compiler to accept non-normalized programs. This included simplifying and generalizing the formalization of clocks presented in [20]. With P. Jeanmaire (M2 internship), we have implemented a compilation pass to translate normalized programs from one syntactical form to another. The main challenge was to formally prove an alignment property (signals are present iff their clocks are true) that had been assumed until now. The proof is finished except for the inductive case for the reset construct which we hope to complete soon.

6.3.3. Strengthening the correctness theorem.

The current correctness theorem assumes that an accepted program can be given a semantics in terms of the mechanized model. It should be possible to prove this fact for programs that pass the initial type-checking and clock-checking algorithms, that can be scheduled, and which never invoke an undefined operation (such as a division by zero). We made good progress on this problem by defining an interpreter for normalized Lustre programs and showing that the results it calculates satisfy the semantic predicates. This initial work gives some useful insights into how to proceed. We presented it at the Synchron 2019 workshop.

6.4. Specifying multi-clock Lustre programs

Participants: Timothy Bourke, Guillaume Iooss, Baptiste Pauget, Marc Pouzet.

It is sometimes desirable to compile a single synchronous language program into multiple tasks for execution by a real-time operating system. We have been investigating this question from three different perspectives.

6.4.1. Harmonic clocks

We studied the extension of a synchronous language with periodic harmonic clocks based on the work of Mandel et al. [31], [37], [32], [35], [36] on n-synchrony and the extension proposed by Forget et al. [33]

Mandel et al. considered a language with periodic clocks expressed as ultimately periodic binary sequences. The decision procedures (equality, inclusion, precedence) for such an expressive language can be very costly. It is thus sometimes useful to apply an envelope-based abstraction, that is, one where sets of clocks are represented by a rational slope and an interval. Forget considered simpler "harmonic" clocks. His decision procedures conincide with those for the envelope-based abstraction but without any loss of information. During his M2 ineternship, B. Pauget continued this line of work by extending the input language of the Vélus Lustre compiler with harmonic clocks. This work was the starting point for the proposal of a new intermediate language for a synchronous compiler that is capable of exploiting clock information to apply agressive optimizations and generate parallel code.

6.4.2. New Intermediate Language MObc (Multi Object Code)

This intermediate language is reminiscent of the intermediate Obc language used in the Vélus and Heptagon compiler, but with some important differences and new features. MObc permits a synchronous function to be represented as a set of named state variables and possibly nested blocks with a partial ordering which express the way blocks can and must be called. In comparison, Obc represents a synchronous function as a set of state variables and a transition function that is itself written in a sequential language. Each block comprises a set of equations in Single Static Assignment (SSA) form, that is, exactly one equation per variable, so as to simplify the implementation of a number of classic optimizations (for example, constant propagation, inlining, common sub-expression elimination, code specialisation). Then, every block is translated into a step function (e.g., a C function). This intermediate language has been designed to facilitate the generation of code for a real-time OS and a multi-core target. This work exploits two older results: the article of Caspi et al. [30] that introduces an object representation for synchronous nodes and a "scheduling policy" that specifies how their methods may be called, and; the work of Pouzet et al. [38] on the calculation of input/output relations to merge calculations. We are preparing and article on this subject.

6.4.3. Clocking constraints, communication latencies, and constraint solving

In this approach, the top-level node of a Lustre program is distinguished from inner nodes. It may contain special annotations to specify the triggering and other details of node instances from which separate "tasks" are to be generated. Special operators are introduced to describe the buffering between top-level instances. Notably, different forms of the `fby` and `current` operators are provided. Some of the operators are under-specified and a constraint solver is used to determine their exact meaning, that is, whether the signal is delayed by zero, one, or more cycles of the receiving clock, which depends on the scheduling of the source and destination nodes. Scheduling is formalized as a constraint solving problem based on latency constraints between some pairs of input/outputs that are specified by the designer. G. Iooss has been prototyping these ideas in the academic Heptagon compiler.

This work is funded by a direct industrial contract with Airbus. In collaboration (this year) with Michel Angot Vincent Bregeon Jean Souyris (Airbus, R&D) and Matthieu Boitrel (Airbus BE).

6.5. The Zelus Language

Participants: Timothy Bourke, Ismail Lakhim-Bennani, Marc Pouzet.

Zelus is our laboratory to experiment our research on programming languages for hybrid systems. It is devoted to the design and implementation of systems that may mix discrete-time/continuous-time signals and systems between those signals. It is essentially a synchronous language reminiscent of Lustre and Lucid Synchrone but with the ability to define functions that manipulate continuous-time signals defined by Ordinary Differential Equations (ODEs). The language is functional in the sense that a system is a function from signals to signals (not a relation). It provides some features from ML languages like higher-order and parametric polymorphism as well as dedicated static analyses.

This year, we have pursued our work on the design, semantics and implementation of hybrid modeling language, in particular the treatment of Differential Algebraic Equations (DAEs) [23].

6.5.1. Compiler Internals: Static Typing and Compiler Organisation

The distribution with manual and examples is distributed at http://zelus.di.ens.fr (only Version 1, in binary form). Version 2 (the current active branch) is available in source form on Inria GitLab https://gitlab.inria.fr/parkas/zelus, on simple demand.

Several new experimentations have been done this year, in particular on the type system and an extensive rewritting of some compilation internals to simplify the code and make the generated code more shorter (in size) and more efficient.

6.5.2. Co-simulation as Function Lifting

Hybrid models in Zelus (that is, programs that mix discrete and continuous-time signals) are simulated using a single ODE and zero-crossing solvers only. All hybrid modeling languages (e.g., Simulink, Modelica, Ptolemy) act the same way, at least, single solver simulation is the default mechanism.

Its weaknesses are well known: any change of the dynamic, even local, calls for a global reset of the solver, making it slower for that later steps; the mix of a slow and fast signals slows down the whole simulation. Co-simulation is about running several solvers (or instances of the same) at the same time.

We proposed a limited (but useful) manner, by proving a way to internalize the solver to obtain, from a continuous-time function, a synchronous stream function. A preliminary experiment done this year was surprisingly and pleasingly simple to implement in Zelus. It consisted in defining a (higher-order) function *solve* that, given a continuous-time function *f* returns a stream function *solve f*. Given an input stream *x* and an increasing stream of time horizons *h*, *solve f(x, t)* returns the stream of approximated values. This function internalizes the ODE solver and the zero-crossing detection mechanism. The overall model is then a purely discrete-time, synchronous model. In particular, classical synchronization protocols between solvers can be programmed in the language itself, hence benefiting from the static checks that track typing, causality and initialization errors, properties that would be more difficult to ensure if programmed directly in C, for example. We think that it is even possible to write a formal synchronous specification of the simulation engine itself, that is, to program the function solve directly in Zelus. This experiment on co-simulation gives new insight on the semantics based on non standard analysis that we proposed and, more interestingly, to relate it to the proven and more classical semantics based on super-dense time studied and exploited by Edward Lee.

6.5.3. QSS-based Simulation

Quantized State Systems simulation (QSS) was introduced in the early 2000's by F. Cellier and E. Kofman as an alternative to time-based simulation, which is the dominant approach to ODE/DAE systems simulation.

Rather than linking QSS to Discrete Event Simulation, we have made a preliminary experiment to relate it to Synchronous Programming and its continuous time extension Zelus. Zelus is used to give a formal description of the QSS method that can be executed. We have described the very basic scheme called QSS (or QSS1) for which we can give a Zelus (hence executable) specification. Higher order schemes QSS2, 3, etc. can also be given an Zelus specification. Implicit schemes were also proposed by Kofman for a better handling of stiff systems. Higher order versions of BQSS are nontrivial; they are called LIQSS1, 2, 3, etc. and they can also be specified in Zelus. This preliminary work is done in collaboration with Albert Benveniste (Inria Hycomes, Rennes) and funded by the Modeliscale FUI project.

6.5.4. Property Based Testing of Hybrid Programs

Property-based program testing involves checking an executable specification by running many tests. We build on the work of Georgios Fainekos and Alexandre Donzé, and take inspiration from earlier work by Nicolas Halbwachs, to write a Zélus library of synchronous observers with a quantitative semantics that can be used to specify properties of a system under test. We implemented several optimization algorithms for producing test cases, some of which are gradient-based. To compute the gradients, we use Automatic Differentiation (AD) of the system under test and its specification. Together with François Bidet, we ported the well-known FADBAD++ library for AD written by Ole Stauning in 1997 to OCaml—the target language of Zélus. Our port is called FADBADml and is now released under an Inria license [0] and is available on opam.

6.6. Reactive Probabilistic Programming

Participant: Marc Pouzet.

Synchronous languages were introduced to design and implement real-time embedded systems with a (justified) enphasis on determinacy. Yet, they interact with a physical environment that is partially known and are implemented on architectures subject to failures and noise (e.g., channels, variable communication delays or computation time). Dealing with uncertainties is useful online monitoring, learning, statistical testing or to build simplified models for faster simulation. Actual synchronous and languages provide limited support for modeling the non-deterministic behaviors that are omnipresent in embedded systems.

In 2019, we started a new topic on *reactive probabilistic programming* under the initiative of Guillaume Baudart and Louis Mandel (IBM Research, Watson); in collaboration with Erik Atkinson, Michael Carbin and Benjamin Sherman (MIT). We have designed ProbZelus, an extension of Zelus with probabilistic programming constructs. The language makes it possible to describe probabilistic models in interaction with an observable environment. At runtime, a set of inference techniques can be used to learn the distributions of model parameters from observed data. The main results are (1) the design of the ProbZelus compiler, the formalization of the static and dynamic semantics of the language that mixes deterministic and probabilistic components, (2) the design and implementation of inference methods which can be executed with bounded resources, (3) the evaluation of ProbZelus on a set of examples and case studies.

For the moment, ProbZelus is mainly a library of Zelus, with minor changes of the language itself (essentially the type system) [0] It exploits heavilly the higher-order nature of Zelus. E.g., if `f` is a stream function (with type `f: 'a -D-> 'b`) and `x` is a stream (with type `'a`), `Particule.infer f x: 'a -D-> 'b Distribution.t` implements an inference algorithm which computes a distribution for the result of type `'b` with a particule filter algorithm.

A prelimary report describes a part of this work [24] and a presentation at JFLA will be given in January 2020. ProbZelus is available in open source at https://github.com/IBM/probzelus since december 2019. Our purpose is to go beyond the library approach with a closer integration of probabilistic constructs and reactive constructs, with dedicated static analyses and compilation techniques to give static guaranties on the result of inference, efficient inference techniques tuned for reactive applications and that ensure execution in bounded time and space; efficient dedicated compilation techniques for probabilistic programs. Finally, the treatment of both discrete-time and continuous-time signals and systems must be investigated (only discrete-time is considered at the moment).

6.7. Identification of matrix operations for Compute-In-Memory architectures from a high-level Machine Learning framework

Participant: Andi Drebes.

Compute-In-Memory (CIM) architectures are capable of performing certain performance-critical operations directly in memory (e.g., matrix multiplications) and represent a promising approach to partially eliminate the bottleneck of traditional von Neumann-based architectures resulting from long-distance communication between main memory and processing units.

In order for applications to benefit from such architectures, their operations must be divided into highly parallel, uniform operations eligible for in-memory computation and control logic that cannot benefit from CIM and that must be carried out by conventional computing devices. It is crucial for this process that as many eligible operations as possible are identified and effectively processed in memory, resulting only in as few computations as possible carried out on the conventional cores.

[0] https://fadbadml-dev.github.io/FADBADml/

[0] Yet, higher-order was only used occasionally since then; hence an important implementation effort has been spent this year to make it work well.

The programmability of CIM architectures is a key factor for its overall success. Manual identification of eligible operations and mapping to hardware resources is tedious, error-prone and requires detailed knowledge of the target architecture and therefore does not represent a viable approach to program CIM architectures.

With our partners from the MNEMOSENE project, we have developed a compilation toolchain that unburdens programmers from technical details of CIM architectures by allowing them to express algorithms at a high level of abstraction and that automates parallelization, orchestration and the mapping of operations to the CIM architecture. The solution integrates the Loop Tactics [40] declarative polyhedral pattern recognition and transformation framework into Tensor Comprehensions [39], a framework generating highly optimized kernels for accelerators from an abstract, mathematical notation for tensor operations. The compilation flow performs a set of dedicated optimizations aiming at enabling the reliable detection of computational patterns and their efficient mapping to CIM accelerators.

The results of this work have been submitted to the 10th International Workshop on Polyhedral Compilation Techniques (IMPACT).

6.8. Applying reinforcement learning to improve a branch-and-bound optimizing compiler

Participant: Basile Clement.

Frameworks for image processing, deep learning, etc., work with Directed Acyclic Graphs (DAGs) of computational operators that wrap high-performance libraries. The production of highly optimized, target-specific implementations of the library functions come at a high engineering cost: languages and compilers have failed to deliver performances competitive with expert written code, notably on GPUs and other hardware accelerators. Moreover, library implementations may not offer optimal performance for a specific use case. They may lack inter-operator optimizations and specializations to specific data sizes and shapes.

In his thesis, Ulysse Beaugnon, a former PhD student in the team, proposed to formulate this compilation problem as an optimization research problem using a combination of analytical modeling, experimental search, constraint programming and branch-and-bound optimization techniques. Basile Clement started a PhD to extend this idea, exploring the improvements required to make it fully competitive with handwritten code. In 2019, he evaluated reinforcement learning techniques such as multi-armed bandit schemes to improve the performance and efficiency of the search procedure; extended the analytical model with generic sizes, making it more precise before selecting tiling parameters; and made various improvements to the code generation procedure.

SPADES Project-Team

6. New Results

6.1. Design and Programming Models

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

6.1.1. Hypercells

The location graph framework we have introduced in [66] has evolved into the Hypercell framework presented in [18]. The Hypercell framework allows the definition of different component models for dynamic software architectures featuring both sharing and encapsulation. The basic behavioral theory of hypercells in the form of a contextual bisimulation has been developed and we are currently developing proofs of correctness for encapsulation policies based on this theory.

In collaboration with the Spirals team at Inria Lille – Nord Europe, and Orange, we have used hypercells as a pivot model for developing interpretations, formally defined with the Alloy specification language, of various languages and formalisms for the description of software configurations for cloud computing environments. Configuration languages considered include the TOSCA and OCCI standards, as well as the Open Stack Heat Orchestration Template (HOT), Docker Compose, and the Aeolus component model for cloud deployment. This work, developed as part of a bilateral contract with Orange, allowed the development of a verification tool for the correctness of HOT configurations, and helped uncover several flaws in the ETSI NFV standard.

6.1.2. 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) [49], [39], we have written a survey providing a comprehensive description of the existing parametric dataflow MoCs [42], and we have studied *symbolic* analyses of dataflow graphs [43]. More recently, we have proposed an original method to deal with lossy communication channels in dataflow graphs [48].

We are nowadays studying models allowing *dynamic reconfigurations* of the *topology* of the dataflow graphs. This is required by many modern streaming applications that 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) [13]. 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. Transformations can be seen as graph rewriting rules that match some sub-part of the dataflow graph and replace it by another one. Transformations can be applied an arbitrary number of times during execution and therefore can produce an arbitrary number of new graphs. The major feature and advantage of RDF is that it can be statically analyzed to guarantee that all possible graphs generated at runtime will be connected, consistent, and live. To the best of our knowledge, RDF is the only dataflow MoC allowing an arbitrary number of topological reconfigurations while remaining statically analyzable. It remains to complete the RDF implementation and to evaluate it on realistic case studies. Preliminary results indicate that dynamic reconfigurations can be implemented efficiently.

This is the research topic of Arash Shafiei's PhD, in collaboration with Orange Labs.

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 [46]. 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 [69]. 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 [35].

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 parallel programming language on PREcision Timed (hence deterministic) architectures [71], [72]. The originality resides in the time-triggered model of computation and communication that allows for a very precise control over the thread execution. Synchronization 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 [17]. This is a collaboration with Nicolas Hili and Eric Jenn, the postdoc of Nicolas Hili being funded by the CAPHCA project.

We have also proposed a *multi-rate* extension of FOREC [16]. Indeed, up to now FOREC programs were constrained to operate at a single rate, meaning that all the parallel threads had to share the same execution rate. While this simplified the semantics, it also represented a significant limitation.

Finally, we have extended the compiler of the PRET-C programming language [33], [34] in order to make it energy aware. PRET-C is a parallel programming language in the same sense as Esterel [44], meaning that the parallelism is "compiled away": the PRET-C compiler generates sequential code where the parallel threads from the source program are interleaved according to the synchronous semantics, and produces a classical Control Flow Graph (CFG). This CFG is then turned into a Timed Control Flow Graph (TCFG) by labeling each basic block with the number of clock cycles required to execute it on the chosen processor, based on its micro-architectural characteristics. From the TCFG, we use the method described in Section 6.2.5 to compute a Pareto front of non-dominated (worst-case execution time – WCET, worst-case energy consumption – WCEC) compromises.

6.2.2. Synthesis of switching controllers using approximately bisimilar multiscale abstractions

The use of discrete abstractions for continuous dynamics has become standard in hybrid systems design (see *e.g.*, [67] 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 [64]. 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 [50]. 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.

In previous work we have proposed an approach using mode sequences as symbolic states for our abstractions [59]. 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 [12]. We have shown the effectiveness of the approach on examples inspired by road traffic regulation.

6.2.3. 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 minimize the energy consumption while guaranteeing that each job meets its deadline. The idea is to leverage on the *statistical information* on the jobs' characteristics available at design time: release time, worst-case execution time (WCET), and relative deadline. This is the topic of Stephan Plassart's PhD, funded by the CASERM Persyval project. We have considered several cases depending on the amount of information available at design time:

Offline case: In the offline case, all the information is known and we have proposed the first linear complexity offline scheduling algorithm that minimizes the total energy consumption [15]: our complexity is $\mathcal{O}(n)$ where n is the number of jobs to be scheduled, while the previously best known algorithms were in $\mathcal{O}(n^2)$ and $\mathcal{O}(n \log n)$ [60].

Clairvoyant case: In the clairvoyant case, the characteristics of the jobs are only known statistically, and each job's WCET and relative deadline are only known at release time. We want to compute the *optimal* online scheduling speed policy that minimizes the *expected* energy consumption while guaranteeing that each job meets its deadline. 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 the finite horizon case we use a dynamic programming algorithm, while in the infinite horizon case we use a value iteration algorithm [25].

Non-clairvoyant case: In the non-clairvoyant case, the actual execution time (AET) of a job is only known only when this job completes its execution. This AET is of course assumed to be less than the WCET, which is known at the job's release time. Again, by building an MDP for the system with a well chosen state, we compute the *optimal* online scheduling speed policy that minimizes the *expected* energy consumption [26].

Learning case: In the learning case, the only information known for the jobs are a bound on the jobs' WCETs and a bound on their deadlines. We have proposed two *reinforcement learning* algorithms, one that learns the optimal value of the expected energy (Q-learning), and another one that learns the probability transition matrix of the system, from which we derive the optimal online speed policy.

This work led us to compare several existing speed policies with respect to their feasibility. Indeed, the policies (OA) [70], (AVR) [70], and (BKP) [37] 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 [27].

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

We contribute to Prosa [31], 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.

We have further developed CertiCAN, a tool produced using Coq for the formal certification of CAN analysis results [14]. 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 [68]. Additional optimizations have been implemented (and proved correct) to make CertiCAN computationally efficient. Experiments demonstrate that CertiCAN is able to certify the results of RTaW-Pegase, an industrial CAN analysis tool, even for large systems.

In addition, we have started investigating how to connect Prosa with implementations and less abstract models. Specifically, we have used Prosa to provide a schedulability analysis proof for RT-CertiKOS, a single-core sequential real-time OS kernel verified in Coq [20]. A connection with a timed-automata based formalization of the CAN specification is also in progress. Our objective with this line of research is to understand and bridge the gap between the abstract models used for real-time systems analysis and actual real-time systems implementation.

Finally, we contributed to a major refactoring of the Prosa library to make it more easily extendable and usable.

6.2.5. Scheduling under multiple constraints and Pareto optimization

We have completed a major work on embedded systems subject to multiple non-functional constraints, by proposing the first of its kind multi-criteria scheduling heuristics for a DAG of tasks onto an homogeneous multi-core chip [9], [23]. 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, namely DAGs of at most 8 tasks. Comparisons showed that the schedules produced by ERPOT are on average only 9% 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.

In a related line of work, we have considered the bi-criteria minimization problem in the (worst-case execution time – WCET, worst-case energy consumption – WCEC) space for real-time programs. To the best of our knowledge, this is the first contribution of this kind in the literature.

A real-time program is abstracted as a Timed Control Flow Graph (TCFG), where each basic block is labeled with the number of clock cycles required to execute it on the chosen processor at the nominal frequency. This timing information can be obtained, for instance, with WCET analysis tools. The target processor is equipped with dynamic voltage and frequency scaling (DVFS) and offers several (frequency f, voltage V) operating points. The goal is to compute a set of non-dominated points in the (WCET, WCEC) plane, non-dominated in the Pareto sense. Each such point is an assignment from the set of basic blocks of the TCFG to the set of available (f,V) pairs.

From the TCFG we extract the longest execution path, therefore deriving the WCET and the WCEC for a chosen fixed (f,V) pair. By construction, all the other execution paths are shorter, so this WCET and this WCEC hold for the whole program. This ensures that each single-frequency assignment is a non-dominated point. Then, we study two frequencies assignments, still for the longest execution path. When the frequency switching costs in time and in energy are assumed to be negligible, we demonstrate that each two frequencies (say with f_i and f_j) assignment is a point in the segment between the single frequency assignment at f_i and the single frequency assignment at f_j. We also propose a linear time heuristic to assign a (f, V) pair to all the other blocks (*i.e.*, those not belonging to the longest path) such that all the other execution paths have a shorter WCET and a lesser WCEC. A key result is that we demonstrate that any two frequencies assignment where the two frequencies are not contiguous is dominated either by a single frequency assignment or by a two frequencies assignment with contiguous frequencies. A corollary is that the Pareto front is a continuous piece-wise affine function. Finally, we generalize these results to the case where the frequency switching costs are not negligible. This is the topic of Jia Jie Wang's postdoc.

We evaluate our method and heuristic on a set of hard real time benchmark programs and we show that they perform extremely well. Our DVFS assignment algorithm can also be used as a back-end for the compiler of the PRET-C programming language [33], [34] in order to make it energy aware, thanks to the ability of this compiler to generate TCFGs (see Section 6.2.1).

6.3. Fault Management and Causal Analysis

Participants: Gregor Goessler, Jean-Bernard Stefani, Sihem Cherrared, Thomas Mari, Martin Vassor.

6.3.1. Fault Ascription in Concurrent Systems

Fault ascription is a precise form of fault diagnosis that relies on counterfactual analysis for pinpointing the causes of system failures. Research on counterfactual causality has been marked, until today, by a succession of definitions of causation that are informally validated against human intuition on mostly simple examples. This approach suffers from its dependence on the tiny number and incompleteness of examples in the literature, and from the lack of objective correctness criteria [52].

We have defined in [28] a set of expected properties for counterfactual analysis, and presented a refined analysis that conforms to our requirements. As an early study of the behavior of our analysis under abstraction we have established its monotony under refinement.

6.3.2. Causal Explanations in Discrete Event Systems

Model-Based Diagnosis of discrete event systems (DES) usually aims at detecting failures and isolating faulty event occurrences based on a behavioural model of the system and an observable execution log. The strength of a diagnostic process is to determine *what* happened that is consistent with the observations. In order to go a step further and explain *why* the observed outcome occurred, we borrow techniques from causal analysis.

In [21] we have presented two constructions of explanations that are able to extract the relevant part of a property violation that can be understood by a human operator. Both support partial observability of events. The first construction is based on minimal sub-sequences of the traces of the log that entail a violation of the property. The second approach is based on a construction of layers similar to [56], in which the explanation is constructed from the choices that definitely move the system closer to the violation of the property. Both approaches are complementary: while subsequence-based explanations are well suited to "condense" the execution trace in sequential portions of the model but are prone to keep non-pertinent parts such as initialisation sequences in the explanation, effective choice explanations highlight the "fateful" choices in an execution, as well as alternative events that would have helped avoid the outcome. Effective choice explanations are therefore able to explain failures stemming from non-deterministic choices, such as concurrency bugs.

6.3.3. Fault Management in Virtualized Networks

From a more applied point of view we have been 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.

In [29] we have proposed a model-based root cause analysis framework called SAKURA. In order to overcome the lack of accurate previous knowledge, SAKURA features a self-modeling algorithm that models the dependencies within and between layers of virtual networks, including auto-recovery and elasticity aspects. Model-based diagnosis is performed using constraint solving on the previous and acquired knowledge. As an illustration we have applied SAKURA to the virtual IpMultimedia Subsystem (vIMS).

Finally, in our survey on fault management in network virtualization environments [11] we have addressed the impact of virtualization on fault management, proposed a new classification of the recent fault management research achievements in network virtualization environments, and compared their major contributions and shortcomings.

TEA Project-Team

7. New Results

7.1. ADFG: Affine data-flow graphs scheduler synthesis

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

ADFG (Affine DataFlow Graph) synthesizes scheduling parameters for real-time systems modeled as synchronous data flow (SDF), cyclo-static dataflow (CSDF), and ultimately cyclo-static dataflow (UCSDF) graphs. It aims at mitigating the 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. ADFG was originally implemented by Adnan Bouakaz [0]. It is now being collaboratively developed with team Tea, Hai Nam Tran (UBO) Alexandre Honorat (INSA) and Shuvra Bhattacharyya (UMD/INSA/Inria).

ADFG is extended to support automated code generation of the computed buffer sizes and scheduling parameters for dataflow applications that are implemented in the Lightweight Dataflow Environment (LIDE) [0]. LIDE is a flexible, lightweight design environment that allows designers to experiment with dataflow-based implementations directly. LIDE actors and buffers (FIFOs) can be initialized with parameters, including buffer sizes. The usage of LIDE allows a systematic way to instantiate dataflow graphs with the buffer size parameters computed by ADFG.

Actor models and scheduling algorithms in ADFG have been 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 existing 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. Parallel Composition and Modular Verification of Computer Controlled Systems in Differential Dynamic Logic

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

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

[0] S. Lin, Y. Liu, K. Lee, L. Li, W. Plishker, and S. S. Bhattacharyya. 2017. The DSPCAD framework for modeling and synthesis of signal processing systems. Handbook of Hardware/Software Codesign (2017), 1185–1219.

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 the decomposition of systems into components enhanced with compositional 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 particularly adapted to 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 in industry.

Computer-Controlled Systems (CCS) are a subclass of hybrid systems where the periodic relation of control components to time is of paramount importance. Since they additionally are at the heart of many safety-critical devices, it is of primary importance to correctly model such systems and to ensure they function correctly according to safety requirements. Differential dynamic logic $d\mathcal{L}$ is a powerful logic to model hybrid systems and to prove their correctness. We contributed a compositional modeling and reasoning framework to $d\mathcal{L}$ that separates models into components with timing guarantees, such as reactivity of controllers and controllability of continuous dynamics. Components operate in parallel, with coarse-grained interleaving, periodic execution and communication. We present techniques to automate system safety proofs from isolated, modular, and possibly mechanized proofs of component properties parameterized with timing characteristics.

7.3. Multithreaded code generation for process networks

Participants: Loïc Besnard, Thierry Gautier.

As part of an in-depth comparison of process models, we have recently revisited the relation between the model of asynchronous dataflow represented by Kahn Process Networks (KPNs) and that of synchronous dataflow represented by the polychronous model of computation. In particular, we have precisely described in which conditions polychronous programs can be seen as KPNs. In this context, we have considered different cases of process networks, including so-called “polyendochronous processes”. Under some conditions expressed by clock equation systems, (networks of) processes exhibiting polyhierarchies of clocks are polyendochronous and, as compositions of endochronous processes, may be seen as KPNs.

Based on this characterization, we have developed in the open-source Polychrony toolset a new strategy of code generation for such (polyendochronous) process networks. Typically, after the clock calculus, a program P is organized as a composition of processes, $P = (| \ P1 \mid P2 \mid ... \mid Pn \ |)$, each one structured around a clock tree. When P is characterized as polyendochronous, it contains generally clock constraints such as $\mathit{Clk}1 = \mathit{Clk}2$, with $\mathit{Clk}1$ being a clock in the subtree corresponding to $P1$ and $\mathit{Clk}2$ a clock in the subtree corresponding to $P2$.

Such a constraint induces a synchronization between two parts ($P1$, $P2$) of the program when $\mathit{Clk}1$ or $\mathit{Clk}2$ occurs. The principle of the code generation for polyendochronous processes is based on the existing distributed code generation, but with the additional resynchronization of parts of the application induced by the constraints on clocks ($\mathit{Clk}1$, $\mathit{Clk}2$) not placed in the same clock trees. For distributed code generation, it is considered that each (clock) hierarchy will run on a specific processor. In this case, the purpose is mainly to partition the application, and the processors will be virtual ones.

The code generation of each partition consists in the definition of several tasks: one task per cluster (a cluster being a subpart that may be executed as soon as its inputs are available, without any communication with the external world); one task per input/output of the partition; one task for the cluster of state variables; one task that manages the steps. Synchronization between these tasks is obtained by semaphores (one semaphore per task). This code generation technique for the class of networks called “polyendochronous processes” has been

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

added in the Polychrony toolset (http://polychrony.inria.fr) and a paper describing the comparison of process models is currently in submission.

7.4. Type theory for modular static analysis of system programs

Participants: Lucas Franceschino, Jean-Pierre Talpin, David Pichardie.

This Ph.D. project is about formal verification, with system programming applications in mind. Formal methods are essential for safety-critical software (i.e. transport and aeronautic industry). In the same time, more and more programming languages with a strong type system arise (such as Haskell, Rust, ML, Coq, F*, Idris...).

Formal methods come in different flavors: type theory, abstract interpretation, refinement types. Each of these "flavors" are both theoretical fields and are also being implemented concretely: *Astrée* ou *Verasco* for abstract interpretation, *Coq*, *Agda*, *F** or *Idris* dependent types, and *Liquid Haskell* for refinement types.

Our approach consists in positioning ourselves between type theory and abstract interpretation, and to leverage the power of both. The main intuition behind this idea is that abstract interpretation, suffering from expressiveness, would bring *invariant inference* power, while strong type systems, requiring manual annotations and proofs, would bring *expressivity*.

We formalized how one can enrich a weakest precondition calculus (WP) with an abstract interpreter. This work takes the shape of a WP calculus transformer: given a WP calculus, we generically construct a brand new WP calculus that produces easier (but sound, still) weakest preconditions, thanks to abstract interpretation.

Concretely, our work is being implemented as an F* effect transformer that leverage Verasco capabilities, for a low-level subset of F*, namely Low*.

7.5. Verified information flow of embedded programs

Participants: Jean-Joseph Marty, Lucas Franceschino, Niki Vazou, Jean-Pierre Talpin.

This PhD project is about applying refinement types theory to verified programming of applications and modules of library operating systems, such as unikernels, for embedded devices of the Internet of Things (IoT): TinyOS, Riot, etc. Our topic has focused on developing a model of information flow control using labeled input-outputs (LIO) implemented using F☆: project Lio☆.

As part of the development of Lio☆, we implemented a library that, thanks to static verification, ensures the containment of information in relation to a parameterized policy for information flow control. In collaboration with Niki Vazou (IMDEA) and Lucas Franceschino we have formalized and developed an automatic method to prove non-interference in Meta☆. Using the Kremlin code generator, programs using Lio☆ can be compiled into C code and run natively on embedded low-resource-constrained devices, without the need for additional runtime system.

In parallel we continued our collaboration with the ProgSys team on a second, now discontinued, project: Gluco☆. The goal of this project was to evaluate the capabilities to use the F* programming language to program an entire system by taking into account its software, hardware and physical constraints using type refinements [0].

[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, 2019.

ANTIQUE Project-Team

7. New Results

7.1. Relational Static Analysis

7.1.1. Relational abstraction for memory properties

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

Static analyses aim at inferring semantic properties of programs. We can distinguish two important classes of static analyses: state analyses and relational analyses. While state analyses aim at computing an over-approximation of reachable states of programs, relational analyses aim at computing functional properties over the input-output states of programs. Several advantages of relational analyses are their ability to analyze incomplete programs, such as libraries or classes, but also to make the analysis modular, using input-output relations as composable summaries for procedures. In the case of numerical programs, several analyses have been proposed that utilize relational numerical abstract domains to describe relations. On the other hand, designing abstractions for relations over input-output memory states and taking shapes into account is challenging. We have proposed a set of novel logical connectives to describe such relations, which are inspired by separation logic. This logic can express that certain memory areas are unchanged, freshly allocated, or freed, or that only part of the memory was modified. Using these connectives, we have built an abstract domain and design a static analysis that over-approximates relations over memory states containing inductive structures. We implemented this analysis and evaluated it on a basic library of list manipulating functions.

This work was done as part of the Phd of Hugo Illous [10] and a journal paper is currently under submission.

7.2. Static Analysis of Probabilistic Programming Languages

7.2.1. Towards the verification of semantic assumptions required by probabilistic inference algorithms

Participants: Wonyeol Lee, Hangyeol Wu, Xavier Rival [correspondant], Hongseok Yang.

Probabilistic programming is the idea of writing models from statistics and machine learning using program notations and reasoning about these models using generic inference engines. Recently its combination with deep learning has been explored intensely, which led to the development of so called deep probabilistic programming languages, such as Pyro, Edward and ProbTorch. At the core of this development lie inference engines based on stochastic variational inference algorithms. When asked to find information about the posterior distribution of a model written in such a language, these algorithms convert this posterior-inference query into an optimisation problem and solve it approximately by a form of gradient ascent or descent. We analysed one of the most fundamental and versatile variational inference algorithms, called score estimator or REINFORCE, using tools from denotational semantics and program analysis. We formally expressed what this algorithm does on models denoted by programs, and exposed implicit assumptions made by the algorithm on the models. The violation of these assumptions may lead to an undefined optimisation objective or the loss of convergence guarantee of the optimisation process. We then describe rules for proving these assumptions, which can be automated by static program analyses. Some of our rules use nontrivial facts from continuous mathematics, and let us replace requirements about integrals in the assumptions, such as integrability of functions defined in terms of programs' denotations, by conditions involving differentiation or boundedness, which are much easier to prove automatically (and manually). Following our general methodology, we have developed a static program analysis for the Pyro programming language that aims at discharging the assumption about what we call model-guide support match. Our analysis is applied to the eight representative model-guide pairs from the Pyro webpage, which include sophisticated neural network models such as AIR. It found a bug in one of these cases, and revealed a non-standard use of an inference engine in another, and showed that the assumptions are met in the remaining six cases.

This work has been published in [12].

7.3. Static Analysis of JavaScript Code

7.3.1. Weakly Sensitive Analysis for Unbounded Iteration over JavaScript Objects

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

In [23] and [11], we studied composite object abstraction for the analysis JavaScript.

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

7.4. Rule-based Modeling with Arithmetics

7.4.1. Counters in Kappa: Semantics, Simulation, and Static Analysis.

Participants: Pierre Boutillier, Ioana Cristescu, Jérôme Feret.

Site-graph rewriting languages, such as Kappa or BNGL, offer parsimonious ways to describe highly combinatorial systems of mechanistic interactions among proteins. These systems may be then simulated efficiently. Yet, the modeling mechanisms that involve counting (a number of phosphorylated sites for instance) require an exponential number of rules in Kappa. In BNGL, updating the set of the potential applications of rules in the current state of the system comes down to the sub-graph isomorphism problem (which is NP-complete).

In [14], we extend Kappa to deal both parsimoniously and efficiently with counters. We propose a single push-out semantics for Kappa with counters. We show how to compile Kappa with counters into Kappa without counters (without requiring an exponential number of rules). We design a static analysis, based on affine relationships, to identify the meaning of counters and bound their ranges accordingly.

7.5. Reduced product

7.5.1. Sharing Ghost Variables in a Collection of Abstract Domains.

Participants: Marc Chevalier, Jérôme Feret.

In abstract interpretation, it is often necessary to be able to express complex properties while doing a precise analysis. A way to achieve that is to combine a collection of domains, each handling some kind of properties, using a reduced product. Separating domains allows an easier and more modular implementation, and eases soundness and termination proofs. This way, we can add a domain for any kind of property that is interesting. The reduced product, or an approximation of it, is in charge of refining abstract states, making the analysis precise.

In program verification, ghost variables can be used to ease proofs of properties by storing intermediate values that do not appear directly in the execution.

In [15], we propose a reduced product of abstract domains that allows domains to use ghost variables to ease the representation of their internal state. Domains must be totally agnostic with respect to other existing domains. In particular the handling of ghost variables must be entirely decentralized while still ensuring soundness and termination of the analysis.

7.6. Static Analysis of Neural Networks

7.6.1. Perfectly Parallel Fairness Certification.

Participants: Caterina Urban [correspondant], Maria Christakis, Valentin Wüestholz, Fuyuan Zhang.

Recently, there is growing concern that machine-learning models, which currently assist or even automate decision making, reproduce, and in the worst case reinforce, bias of the training data. The development of tools and techniques for certifying fairness of these models or describing their biased behavior is, therefore, critical.

In [19], we propose a perfectly parallel static analysis for certifying causal fairness of feed-forward neural networks used for classification tasks. When certification succeeds, our approach provides definite guarantees, otherwise, it describes and quantifies the biased behavior. We design the analysis to be sound, in practice also exact, and configurable in terms of scalability and precision, thereby enabling pay-as-you-go certification. We implement our approach in an open-source tool and demonstrate its effectiveness on models trained with popular datasets.

7.7. Reductions between synchronous and asynchronous programming abstractions

7.7.1. Communication closed asynchronous protocols.

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

Fault-tolerant distributed systems are implemented over asynchronous networks, where performance emerges from the load of the system. 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 [17], we defined the first algorithm that automatically transforms an asynchronous protocol into a synchronous one. The method is sound but not complete. The transformation is based on an axiomatization of the notion of communication closure introduce by Elrad and Frances. If the asynchronous protocol is communication-closed then the translator will successfully compute its synchronous counter-part. Checking communication closure is done locally without considering any interferences between processes. The translator was successfully applied to Multi-Paxos, ViewStamped, and the atomic broadcast of Chandra and Toueg, generating the first synchronous counterparts of these protocols. The transformation from asynchronous to synchronous preserves the local states process go through and the exchanged messages. The translator has been implemented in a prototype tool called Athos, i.e., Asynchronous To Heard-Of Synchronizer, that is open source. The tool takes as input protocols in an intermediate protocol languages that has an asynchronous semantics and it is very close to C. These results have been published in one of the main verification venues Computer Aided Verification, CAV 2019 (acceptance rate <25% out of >250 submissions). The impact of the translator from asynchronous protocols to equivalent synchronous ones is important for the verification community because such a transformation reduces dramatically the state space and the set of traces to explore in order to prove the program correct, independently of the used verification technique.

7.7.2. Executable Rounds: a Programming Abstraction for Fault-Tolerant Protocols.

Participants: Cezara Drăgoi, Josef Widder, Damien Zufferey.

Fault-tolerant distributed systems are notoriously difficult to design and implement. Although programming languages for distributed systems is an active research area, appropriate synchronization primitives for fault-tolerance and group communication remains an important challenge. In [18] we present a new programming abstraction, HSync, for implementing benign and Byzantine distributed protocols. HSync is based on communication-closed rounds. Round models offer a simple abstraction for group communication and communication-closed rounds simplify dealing with faults. Protocols are implemented in a modular way in HSync. The language separates the message reception from the process local computation. It extends classic rounds with language constructs that give to the programmer the possibility to implement network and algorithm-specific policies for message reception. We have implemented an execution platform for HSync that runs on top of commodity hardware. We evaluate experimentally its performance, by comparing consensus implementations in HSync with LibPaxos3 and Bft-SMaRt, two consensus libraries tolerant to benign, resp. Byzantine faults.

7.8. Introduction

Participant: Andreea Beica.

The PhD of Andreea Beica [9] aims at studying two aspects related to the modelling of Biochemical Reaction Net- works, in the context of Systems Biology.

In the first part, we analyse how scale-separation in biological systems can be exploited for model reduction. We first argue for the use of rule-based models for prototyping genetic circuits, and then show how the inherent multi-scaleness of such systems can be used to devise a general model approximation method for rule-based models of genetic regulatory networks. The reduction proceeds via static analysis of the rule system. Our method relies on solid physical justifications, however not unlike other scale-separation reduction techniques, it lacks precise methods for quantifying the approximation error, while avoiding to solve the original model. Consequently, we next propose an approximation method for deterministic models of biochemical networks, in which reduction guarantees represent the major requirement. This second method combines abstraction and numerical approximation, and aims at providing a better understanding of model reduction methods that are based on time- and concentration- scale separation.

In the second part of the thesis, we introduce a new re-parametrisation technique for differential equation models of biochemical networks, in order to study the effect of intracellular resource storage strategies on growth, in self-replicating mechanistic models. Finally, we aim towards the characterisation of cellular growth as an emergent property of a novel Petri Net model semantics of Biochemical Reaction Networks.

CAMBIUM Project-Team

6. New Results

6.1. Programming language design and implementation

6.1.1. The OCaml system

Participants: Damien Doligez, Armaël Guéneau, Xavier Leroy, Luc Maranget, David Allsop [Cambridge University], Florian Angeletti, Frédéric Bour [Facebook, until Sep 2019], Stephen Dolan [Cambridge University], Alain Frisch [Lexifi], Jacques Garrigue [Nagoya University], Sébastien Hinderer [SED], Nicolás Ojeda Bär [Lexifi], Gabriel Radanne, Thomas Refis [Jane Street], Gabriel Scherer [Inria team Parsifal], Mark Shinwell [Jane Street], Leo White [Jane Street], Jeremy Yallop [Cambridge University].

This year, we released four versions of the OCaml system: versions 4.08.0, 4.08.1, 4.09.0, and 4.09.1. Versions 4.08.1 and 4.09.1 are minor releases that respectively fix 6 and 5 issues. Versions 4.08.0 and 4.09.0 are major releases that introduce new language features, improve performance and usability, and fix about 50 issues. The main novelties are:

- User-defined binding operators are now supported, with syntax similar to `let*`, `let+`, `and*`. These operators make it much easier to write OCaml code in monadic style or using applicative structures.
- The `open` construct now applies to arbitrary module expressions in structures and to applicative paths in signatures.
- A new notion of user-defined “alerts” generalizes the “deprecated” warning.
- New modules were added to the standard library: `Fun`, `Bool`, `Int`, `Option`, `Result`.
- Many floating-point functions were added, including fused multiply-add, as well as a new `Float.Array` submodule.
- Many error messages were improved, as well as error and warning reporting mechanisms.
- Pattern-matching constructs that correspond to affine functions are now optimized into arithmetic computations.

6.1.2. Evolution of the OCaml type system

Participants: Florian Angeletti, Jacques Garrigue, Thomas Refis [Jane Street], Didier Rémy, Gabriel Radanne, Gabriel Scherer [Inria team Parsifal], Leo White [Jane Street].

In addition to the work done on the above releases, efforts have been done to improve the type system and its implementation. Those include:

- Formalizing the typing of the pattern-matching of generalized algebraic data types (GADTs).
- Fixing some issues related to the incompleteness of the treatment of GADTs.
- Proposing extensions of the type system to reduce this incompleteness in concrete cases, by refining the information on abstract types.
- Exploring practical ways to obtain more polymorphism for functions whose soundness does not rely on the value restriction.
- Improving the readability of the type-checker code.
- Making the module layer of the type-checker more incremental, in order to improve efficiency and to facilitate integration with documentation tools.

6.1.3. Refactoring with ornaments in ML

Participants: Didier Rémy, Thomas Williams [Google Paris].

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.

This year, Williams and Rémy improved the formalization of the lifting framework. In particular, they introduced an intermediate language, in which nonexpansive expressions can be marked on source terms and traced during reduction. This allows to treat the nonexpansive part of expansive expressions as nonexpansive and use equational reasoning on nonexpansive parts of terms that appear in types. This approach significantly simplifies the metatheory of ornaments. This calculus could also have some interest in itself, beyond ornaments, to study languages with side effects.

6.1.4. A better treatment of type abbreviations during type inference

Participants: Didier Rémy, Carine Morel.

During her M2 internship under the supervision of Didier Rémy, Carine Morel revisited the treatment of type abbreviations in type inference for ML-like type systems, using a modern approach based on typing constraints [24]. Instead of expanding type abbreviations prior to unification, both the original abbreviated view and all expanded views are kept during unification, so as to avoid unnecessary expansions and use the least-expanded view whenever possible in the result of unification.

6.2. Software specification and verification

6.2.1. The CompCert formally-verified compiler

Participants: Xavier Leroy, Jacques-Henri Jourdan [CNRS], Michael Schmidt [AbsInt GmbH], Bernhard Schommer [AbsInt GmbH].

In the context of our work on compiler verification, 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 [8]. This compiler comprises a back-end part, which translates the Cminor intermediate language to PowerPC assembly and which is reusable for source languages other than C [7], and a front-end, which translates 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 added a new optimization to CompCert: "if-conversion", that is, the replacement of conditional statements and expressions by conditional move operations and similar branchless instruction sequences. As a consequence, fewer conditional branch instructions are generated. This replacement usually improves worst-case execution time (WCET), because mispredicted conditional branches tremendously increase execution time. This replacement is also interesting for cryptographic code and other programs that manipulate secret data: conditional branches over secret data take time that depends on the data, leaking some information, while conditional move instructions are constant-time and do not leak. The new if-conversion optimization plays a role in the ongoing work of Inria team Celtique on compilation that preserves constant-time properties. Its proof of semantic preservation is nontrivial and prompted the development of a new kind of simulation diagram.

Other recent improvements to the CompCert C compiler include:

- a new code generator targeting the AArch64 instruction set, that is, the 64-bit mode of the ARMv8 architecture;
- the ability to specify the semantics of certain built-in functions, making them amenable to optimizations such as constant propagation and common subexpression elimination;
- improvements to the verified C parser generated by Menhir, including fewer run-time checks, faster validation, and the removal of all axioms from the proof.

We released two versions of CompCert incorporating these improvements: version 3.5 in February 2019 and version 3.6 in September 2019.

6.2.2. Time credits and time receipts in Iris

Participants: Glen Mével, François Pottier, Jacques-Henri Jourdan [CNRS].

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 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 [12] 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 carried out in 2018 has been published at ESOP 2019 [16].

6.2.3. A program logic for Multicore Ocaml

Participants: Glen Mével, François Pottier, Jacques-Henri Jourdan [CNRS].

Glen Mével, who is co-advised by Jacques-Henri Jourdan and François Pottier, has been working on designing a mechanized program logic for Multicore OCaml.

One of the key challenges is to enable deductive reasoning under a weak memory model. In such a model, the behaviors of a program are no longer described by a naive interleaving semantics. Thus, the operational semantics that describes a weak memory model often feels unnatural to the programmer, and is difficult to reason about.

This year, Glen designed and implemented a proof system on top of Iris, a modular separation logic framework whose implementation and soundness proof are both expressed in Coq. This system allows mechanized program verification for a fragment of the Multicore OCaml language. It provides a certain degree of abstraction over the low-level operational semantics, in the hope of simplifying reasoning. This abstraction includes an abstract concept of “local view” of the shared memory; views are exchanged between threads via atomic locations.

A few simple concurrent data structures have been proven correct using the system. They include several variants of locks and mutual exclusion algorithms.

Glen presented preliminary results at the Iris Workshop in October 2019.

6.2.4. Verifying a generic local solver in Iris

Participants: Paulo Emílio de Vilhena, Jacques-Henri Jourdan [CNRS], François Pottier.

From March to August 2019, Paulo Emílio de Vilhena did an M2 internship in our team, where he was advised by François Pottier, with precious help from Jacques-Henri Jourdan (CNRS).

Paulo verified a short but particularly subtle piece of code, namely a "local generic solver", that is, an on-demand, incremental, memoizing least fixed point computation algorithm. This algorithm is a slightly simplified version of `Fix`[0], an OCaml library published by François Pottier in 2009.

[0] https://gitlab.inria.fr/fpottier/fix

The specification of this algorithm is simple: the solver computes the optimal least fixed point of a system of monotone equations. Although the solver relies on mutable internal state for memoization and for "spying", a form of dynamic dependency discovery, no side effects are mentioned in the specification. The challenge is precisely to formally justify why it is permitted to hide these side effects from the user.

The verification is carried out in Iris, a modern breed of concurrent separation logic. Iris is embedded in Coq, so the proof is machine-checked. The proof makes crucial use of prophecy variables, a novel feature of Iris. Auxiliary contributions include a restricted infinitary conjunction rule for Iris and a specification and proof of Longley's "modulus" function, an archetypical example of spying.

This paper [13] has been accepted for presentation at the conference POPL 2020, which will take place in New Orleans in January 2020.

6.2.5. Formal reasoning about asymptotic complexity

Participants: Armaël Guéneau, Arthur Charguéraud [Inria team Camus], François Pottier, Jacques-Henri Jourdan [CNRS].

For several 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 2018 and 2019, in collaboration with Jacques-Henri Jourdan, Armaël has worked on a more ambitious case study, namely a state-of-the-art incremental cycle detection algorithm, whose amortized complexity analysis is nontrivial. Armaël has proposed an improved and simplified algorithm and has carried out a machine-checked proof of its complexity. Furthermore, the verified algorithm has been released and is now used in production inside the Dune build system for OCaml. A paper has been published and presented at the International Conference on Interactive Theorem Proving (ITP 2019) [15]. A more detailed version of these results appears in Armaël Guéneau's dissertation [11], which was defended on December 16, 2019.

6.2.6. TLA+

Participants: Damien Doligez, Leslie Lamport [Microsoft Research], Ioannis Filippidis, Stephan Merz [Inria team VeriDis].

Damien Doligez is the 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 [25], and to build tools for writing TLA+ specifications and mechanically checking the proofs.

We have made a bug-fix release of TLAPS (version 1.4.4). In parallel, we are working on adding features for dealing with temporal properties, that is, fairness and liveness. We have implemented support for the ENABLED operator and the action composition operator in TLA+ proofs. This support is still experimental, but we hope to release a new version of TLAPS next year with these features.

6.3. Shared-memory concurrency

6.3.1. Instruction fetch in the ARMv8 architecture

Participants: Luc Maranget, Peter Sewell [University of Cambridge], Ben Simmer [University of Cambridge].

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.

Luc Maranget participates in project REMS, for *Rigorous Engineering for Mainstream Systems*, an EPSRC project led by Peter Sewell. This year Luc Maranget took part in a research effort that resulted in a paper entitled *ARMv8-A system semantics: instruction fetch in relaxed architectures*. This paper has been accepted for presentation at ESOP 2020. This paper introduces a robust model of instruction fetch and cache maintenance, a central aspect of a processor system's semantics, for ARMv8-A. Luc Maranget specifically extended the **litmus** and **diy** test generators so as to account for self-modifying code. He also performed part of the experiments that support the instruction fetch model.

6.3.2. An ARMv8 mixed-size memory model

Participants: Luc Maranget, Jade Alglave [ARM Ltd & University College London].

Jade Alglave and Luc Maranget have completed their work on a mixed-size version of the ARMv8 memory model. This model builds on the `aarch64.cat` model authored by Will Deacon (ARM Ltd). The model is now ready, and a paper has been written. They hope to work around certain intellectual property restrictions and to submit this paper for publication next year.

6.3.3. Work on diy

Participants: Luc Maranget, Jade Alglave [ARM Ltd & University College London], Antoine Hacquard.

The **diy** suite (for "Do It Yourself") provides a set of tools for testing shared memory models: the **litmus** tool for running tests on hardware, various generators for producing tests from concise specifications, and **herd**, a memory model simulator. Tests are small programs written in x86, Power, ARM, generic (LISA) assembler, or a subset of the C language that can thus be generated from concise specifications, run on hardware, or simulated on top of memory models. Test results can be handled and compared using additional tools. On distinctive feature of our system is Cat, a domain-specific language for memory models.

This year, new synchronisation primitives and instructions were added to various models. Some sizable developments occurred that facilitate the integration of mixed-size models into **herd**: a default definition of the same-instruction relation, which allows using mixed-size models on all tests; an automatic adjustment of the machine's elementary granularity, which facilitates massive testing; and the addition of equivalence classes and relations on them as basic values, which extends the expressiveness of Cat to some abstract mixed-size models.

During a 3-month internship, Antoine Hacquard (an EPITA second-year student) extended the complete tool suite to handle a new target, namely X86_64. The addition of this new target significantly enhances the **diy** tool suite, as X86_64 is a very popular architecture. Moreover, Antoine Hacquard implemented all memory access instructions for all sizes (from byte to quadword), which enabled us to design a mixed-size TSO model for this very popular architecture.

6.3.4. Unifying axiomatic and operational weak memory models

Participants: Quentin Ladeveze, Jean-Marie Madiot, Jade Alglave [ARM Ltd & University College London], 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.

Since October 2019, Luc Maranget and Jean-Marie Madiot are advising a PhD candidate, Quentin Ladeveze. His goal is to further generalize and formalize weak memory models. This involves reasoning about linearizations of interdependent acyclic relations.

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.

CELTIQUE Project-Team

4. New Results

4.1. Compiling Sandboxes:Formally Verified Software Fault Isolation

Participants: Frédéric Besson, Sandrine Blazy, Alexandre Dang, Thomas Jensen.

Software Fault Isolation (SFI) is a security-enhancing pro- gram transformation for instrumenting an untrusted binary module so that it runs inside a dedicated isolated address space, called a sandbox. To ensure that the untrusted module cannot escape its sandbox, existing approaches such as Google's Native Client rely on a binary verifier to check that all memory accesses are within the sandbox. Instead of rely- ing on a posteriori verification, we design, implement and prove correct a program instrumentation phase as part of the formally verified compiler CompCert that enforces a sandboxing security property a priori. This eliminates the need for a binary verifier and, instead, leverages the soundness proof of the compiler to prove the security of the sandbox- ing transformation. The technical contributions are a novel sandboxing transformation that has a well-defined C semantics and which supports arbitrary function pointers, and a formally verified C compiler that im- plements SFI. Experiments show that our formally verified technique is a competitive way of implementing SFI [6].

4.2. Information-Flow Preservation in Compiler Optimisations

Participants: Frédéric Besson, Alexandre Dang, Thomas Jensen.

Correct compilers perform program transformations preserving input/output behaviours of programs. Though mandatory, correctness is not sufficient to prevent program optimisations from introducing information-flow leaks that would make the target program more vulnerable to side-channel at- tacks than the source program. To tackle this problem, we propose a notion of Information-Flow Preserving (IFP) program transformation which ensures that a target program is no more vulnerable to passive side-channel attacks than a source program. To protect against a wide range of attacks, we model an attacker who is granted arbitrary memory accesses for a pre-defined set of observation points. We have proposed a compositional proof principle for proving that a transformation is IFP. Using this principle, we show how a translation validation technique can be used to automatically verify and even close information-flow leaks introduced by standard compiler passes such as dead- store elimination and register allocation. The technique has been experimentally validated on the CompCert C compiler [7].

4.3. Formalization of Higher-Order Process Calculi

Participants: Guillaume Ambal, Alan Schmitt.

Guillaume Amabal and Alan Schmitt, in collaboration with Sergueï Lenglet, 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 extended our previous study that compared locally nameless, De Bruijn indices, and nominal binders with an approach based on higher-order abstract syntax. We have discovered that this approach is not as elegant as in other calculi. A journal version is submitted for publication. The Coq scripts can be found at http://passivation.gforge.inria.fr/hopi/.

4.4. Certified Semantics and Analyses for JavaScript

Participants: Samuel Risbourg, Alan Schmitt.

Alan Schmitt and Samuel Risbourg have continued to develop JSExplain, an interpreter for JavaScript that is as close as possible to the specification. The tool is publicly available at https://github.com/jscert/jsexplain. It was presented to the TC39 committee standardizing JavaScript in December to solicit feedback.

4.5. Skeletal Semantics

Participants: Guillaume Ambal, Nathanael Courant, Thomas Jensen, Adam Khayam, Louis Noizet, Vincent Rebiscoul, Alan Schmitt.

The work on skeletal semantics [5], a modular and formal way to describe semantics or programming languages, has intensified during 2019. We have continued to develop `necro`, a tool to manipulate skeletal semantics and generate interpreters in OCaml, mechanized semantics in Coq, and static analyzers. The code is available online (). Several interns and PhD students are also working on skeletal semantics.

Nathanaël Courant has designed a control-flow analyzer for languages written as skeletal semantics. This work is now extended by Vincent Rebiscoul to certify the analyzer.

Louis Noizet is studying the formalization in Coq of natural semantics from skeletal semantics. To this end, he extended the necro tool to automatically generate a Coq formalization. Louis is also very involved in the maintenance of necro.

Guillaume Ambal is studying the language features that can be captured using skeletal semantics, focusing on concurrency and distribution. In this setting, he is building an approach to automatically derive a small-step semantics form a big-step one.

Adam Khayam is writing a formal semantics of the Hop multitier language, an extension of JavaScript to write web applications. As a first step, he is writing a skeletal semantics of JavaScript to validate that our approach scale for complex and sizable semantics.

4.6. Static analyses for proofs of programs

Participants: Oana Andreescu, Thomas Jensen, Stéphane Lescuyer, Benoît Montagu.

Thomas Jensen together with three industrial research engineers Oana Andreescu, Stéphane Lescuyer, and Benoît Montagu, worked on the development of static analyses that help reduce the manual proof effort that is needed to formally verify programs.

They improved the correlation analysis that Oana Andreescu introduced in her Phd thesis, by designing a novel abstract domain. They verified in Coq its semantic properties, and evaluated their approach on an industrial micro-kernel developed at *Prove&Run*. They showed that the technique could reduce the proof burden by two thirds [1].

4.7. Constant-time verification by compilation and static analysis

Participants: Sandrine Blazy, David Pichardie, Alix Trieu.

To protect their implementations, cryptographers follow a very strict programming discipline called constant-time programming. They avoid branchings controlled by secret data as an attacker could use timing attacks, which are a broad class of side-channel attacks that measure different execution times of a program in order to infer some of its secret values. Several real-world secure C libraries such as NaCl, mbedTLS, or Open Quantum Safe, follow this discipline. We propose an advanced static analysis, based on state-of-the-art techniques from abstract interpretation, to report time leakage during programming. To that purpose, we analyze source C programs and use full context-sensitive and arithmetic-aware alias analyses to track the tainted flows. We give semantic evidences of the correctness of our approach on a core language. We also present a prototype implementation for C programs that is based on the CompCert compiler toolchain and its companion Verasco static analyzer. We present verification results on various real-world constant-time programs and report on a successful verification of a challenging SHA-256 implementation that was out of scope of previous tool-assisted approaches. This work has been published in [4] as an extended version of [12].

CONVECS Project-Team

7. New Results

7.1. New Formal Languages and their Implementations

7.1.1. LOTOS and LNT Specification Languages

Participants: Hubert Garavel, Frédéric Lang, Wendelin Serwe.

LNT [6] [31] 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 § 7.5). 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 2019, a new option `-depend` has been added to the LNT_DEPEND, LNT2LOTOS, and LNT.OPEN tools. LNT_DEPEND now supports the case where the user replaces the predefined LNT modules (e.g., BOOLEAN, NATURAL, etc.) with custom versions. LNT_DEPEND has been made faster and displays better error messages. The LOTOS code generated by LNT2LOTOS for parallel compositions could be semantically incorrect and has been fixed.

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

The version 2.x of TRAIAN that we have been developing for almost 20 years is increasingly difficult to maintain. It consists of 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 2.x only exists in a 32-bit version, and sometimes hits the 4 GB RAM limit when dealing with large compiler specifications, such as those of LNT2LOTOS or EVALUATOR 5.

For this reason, we undertook in 2018 a complete rewrite of TRAIAN (version 3.0) 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.x, 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.x, but should be ultimately capable of bootstrapping itself.

In 2019, we continued the development of TRAIAN 3.0, whose grammar and syntax analysis phase was already almost complete. We fully implemented several static program analysis phases, among which the following:

- binding analysis, which associates a declaration to every identifier occurring in the program (e.g., type, channel, variable, event, etc.)
- typing analysis (including resolution of function name overloading), which associates a type to every expression in the program
- type-productivity and type-finiteness analysis, which check respectively whether a type has at least one value and whether a type has a finite number of values

We also fully implemented the C function generation phase and started to implement the C type generation phase. To avoid problems when switching from TRAIAN 2.x to TRAIAN 3.0, TRAIAN 3.0 generates almost exactly the same code as TRAIAN 2.x. The principal differences concern the numbers used to uniquely identify symbols (variables and functions) in the generated C code, because these are often derived from the syntax tree.

TRAIAN 3.0 is checked regularly against a non-regression test suite consisting of 845 correct and 1545 incorrect programs.

In total, the functionalities that remain to be implemented in TRAIAN 3.0 represent less than 32% of the code of TRAIAN 2.x.

7.1.2. Nested-Unit Petri Nets

Participants: Pierre Bouvier, Hubert Garavel.

Nested-Unit Petri Nets (NUPNs) is a model of computation that can be seen as an upward-compatible extension of P/T nets, which are enriched with structural information on their concurrent and hierarchical structure. Such structural information can easily be produced when NUPNs are generated from higher-level specifications (e.g., process calculi) and allows logarithmic reductions in the number of bits required to represent reachable states, thus enabling verification tools to perform better. For this reason, NUPNs have been so far implemented in thirteen verification tools developed in four countries, and adopted by two international competitions (the Model Checking Contest and the Rigorous Examination of Reactive Systems challenge).

In 2019, a journal article [13] has been published, which formalizes the complete theory of NUPNs.

The development of software tools for NUPNs has steadily progressed. The file format for NUPNs has been enhanced and made more precise; the NUPN_INFO tool has been extended with two new options; the CAESAR.BDD tool as been extended with six new options and its capabilities and efficiency improved in many respects.

We also revisited the problem of decomposing a Petri net into a network of automata, a problem that has been around since the early 70s. We reformulated this problem as the transformation of an ordinary, one-safe Petri net into a unit-safe NUPN. We developed various transformation methods, all of which we implemented in a tool chain that combines NUPN tools with third-party software, such as SAT solvers, SMT solvers, and tools for graph colouring and finding maximal cliques. We performed an extensive evaluation of these methods on a collection of more than 12,000 nets from diverse sources, including nets whose marking graph is too large for being explored exhaustively.

7.1.3. 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 (*Business Process Model and Notation*) standard, and several industrial platforms have been developed for supporting their design, modeling, and simulation.

In collaboration with Francisco Durán (University of Málaga, Spain) and Camilo Rocha (University of Cali, Colombia), we proposed an approach for the modeling and analysis of resource allocation for business processes. Our approach enables the automatic computation of measures for precisely identifying and optimizing the allocation of resources in business processes, including resource usage over time. The proposed analysis, especially suited to support decision-making strategies, is illustrated with a case study of a parcel ordering and delivery by a fleet of drones. This work comprises an encoding of a significant and expressive subset of BPMN in rewriting logic, an executable logic of concurrent change that can naturally deal with states and concurrent computations. The encoding is by itself a formal semantics and interpreter of the BPMN subset that captures all concurrent behavior and thus is used to simulate the concurrent evolution of any business process with a given number of resources and replicas. This work led to two publications, in an international conference [19] and an international journal [12].

7.2. Parallel and Distributed Verification

7.2.1. Debugging of Concurrent Systems using Counterexample Analysis

Participant: 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 a novel approach to improve the usability of model checking by simplifying the comprehension of counterexamples. Our approach takes as input an LTS model and an (unsatisfied) temporal logic property, and operates in four steps. First, all counterexamples for the property are extracted from the model. Second, the model is analyzed to identify the actions that skip from correct to incorrect behaviours (intuitively, these are the most relevant actions from a debugging perspective). Third, using a panel of abstraction techniques, these actions are extracted from the counterexamples. Fourth, 3D visualization techniques are used for highlighting specific regions in the model, where a choice is possible between executing a correct behaviour or falling into an erroneous part of the model, according to the property under analysis. We developed a tool named CLEAR to fully automate our approach, and we applied it on real-world case studies from various application areas for evaluation purposes. This allowed us to identify several patterns corresponding to typical cases of errors (e.g., interleaving, iteration, causality, etc.).

These results led to two publications in international conferences [15] [16] and a publication to appear in an international journal [11].

7.2.2. Eliminating Data Races in Parallel Programs using Model Checking

Participants: Radu Mateescu, Wendelin Serwe.

Parallelization of existing sequential programs to increase their performance and exploit recent multi- and many-core architectures is a challenging but inevitable effort. One increasingly popular parallelization approach is based on OpenMP, which enables the designer to annotate a sequential program with constructs specifying the parallel execution of code blocks. These constructs are then interpreted by the OpenMP compiler and runtime, which assigns blocks to threads running on a parallel architecture. Although this scheme is very flexible and not (very) intrusive, it does not prevent the occurrence of synchronization errors (e.g., deadlocks) or data races on shared variables.

In the framework of the CAPHCA project (see § 9.2.1.1), in collaboration with Eric Jenn and Viet Anh Nguyen (IRT Saint-Exupéry, Toulouse), we proposed an iterative method to assist the OpenMP parallelization by using formal methods and verification. In each iteration, potential data races are identified by applying to the OpenMP program a lockset analysis, which computes the set of shared variables that potentially need to be protected by locks. To avoid the insertion of superfluous locks, an abstract, action-based formal model of the OpenMP program in LNT is extracted and analyzed using the EVALUATOR model checker of CADP. Spurious locks are detected by checking ACTL formulas expressing the absence of concurrent execution of shared variables accesses. This work led to an international publication [28].

7.3. Timed, Probabilistic, and Stochastic Extensions

7.3.1. On-the-fly Model Checking for Extended Regular Probabilistic Operators

Participants: Armen Inants, Radu Mateescu.

Specifying and verifying quantitative properties of concurrent systems requires expressive and user-friendly property languages combining temporal, data-handling, and quantitative aspects. To this aim, we undertook the quantitative analysis of concurrent systems modeled as PTSs (*Probabilistic Transition Systems*), whose actions contain channel names, data values, and probabilities. We proposed a new regular probabilistic operator that extends naturally the Until operators of PCTL (*Probabilistic Computation Tree Logic*) [41], by specifying the probability measure of a path characterized by a generalized regular formula involving arbitrary computations on data values. We devised an on-the-fly model checking method for this new operator, based on a combined local resolution of linear and Boolean equation systems.

In 2019, we continued this work as follows:

- The MCL v4 language was conservatively extended with the new probabilistic operator, leading to a new version MCL v5.
- A new version 5 of the EVALUATOR model checker that handles the MCL v5 language, was added to CADP. EVALUATOR 5 is backward compatible with EVALUATOR 4, to which it adds a new option "`-epsilon`" specifying the precision of floating-point computations. A new version 5 of the MCL_EXPAND tool, the front-end common to the EVALUATOR 3, 4, and 5 model checkers, was added to CADP. This version is upward compatible with the previous one (except for slight changes in some error messages), it corrects a bug and brings some optimizations in the C code generated. Two new manual pages "evaluator5" and "mcl5" have been added.
- For certain probabilistic formulas (e.g., expressing the step-bounded reachability of events), the on-the-fly model checking procedure can be optimized by taking advantage of the possible *query containments*, i.e., implications between instances of the formula with different data parameters. We studied query containment in DHMLR (*Data-based Hennessy-Milner Logic with Recursion*), a parameterized equational formalism used as intermediate language for model checking MCL formulas. Our method consists in detecting, by static analysis, the containment orders present in the DHMLR representation of an MCL formula, and using the information about parameterized Boolean variable implications to improve the convergence of the BES resolution algorithms. We implemented the method in a prototype extension of EVALUATOR 5 and of the CAESAR_SOLVE library for BES resolution, and applied it for verifying probabilistic and also functional properties (e.g., bounded inevitability). The experiments we carried out on self-stabilizing protocols and communication protocols over unreliable channels showed reductions of up to 50% in memory and up to 33% in execution time. This work led to a paper submitted to an international conference.

7.4. Component-Based Architectures for On-the-Fly Verification

7.4.1. Compositional Verification

Participants: Frédéric Lang, Radu Mateescu.

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 2019, in addition to small bug corrections, we updated SVL to support version 5 of EVALUATOR, and we corrected a semantic bug in the expansion of meta-operators of SVL.

In collaboration with Franco Mazzanti (ISTI-CNR, Pisa, Italy), we also used the compositional verification tools of CADP in the framework of the RERS'2019 challenge [0], which consisted in verifying 180 LTL properties and 180 CTL properties on large models of concurrent systems having up to 70 concurrent processes and 234 synchronization actions.

[0] http://rers-challenge.org/2019

We applied to these examples the *maximal hiding* technique [48], which consists in hiding in the model all actions that are not necessary to verify the property. We combined this technique with compositional minimization (using the smart reduction heuristic implemented in SVL) as follows:

- In a first attempt, we used the technique consisting in applying minimization modulo either strong bisimulation or divbranching (divergence-preserving branching) bisimulation, depending on the fragment of the modal μ-calculus to which the formula belongs, as proposed in [48]. This was more efficient than non-compositional verification on large models, but not sufficient to verify all RERS problems successfully.
- We then proposed a refinement of this approach, which consists in (1) partitioning the actions of the system to be verified into so-called strong and weak actions, depending on the formula, and (2) minimizing modulo divbranching bisimulation all processes and process compositions containing weak actions only. This is an improvement over the previous technique, since divbranching bisimulation can be used to minimize some processes of the system even though the formula does not belong to the fragment of the μ-calculus adequate with divbranching bisimulation (which corresponds to formulas with an empty set of strong actions). This new technique allowed us to verify a lot more problems successfully, but still letting a few of the largest RERS problems unresolved. We published a paper describing the approach in an international conference [23].
- At last, we designed a new bisimulation relation, named *sharp bisimulation*, parameterized by the strong actions of the system, and we implemented a prototype tool that reduces a behavior graph modulo this relation. Sharp bisimulation parameterized by a set S of strong actions is weaker than strong bisimulation, stronger than divbranching bisimulation, and adequate with formulas whose strong actions are included in S. Such a fine-tuning of the bisimulation relation by strong actions allowed us to verify all RERS problems successfully and to win the 2019 challenge. A paper describing the approach was accepted for publication in an international conference.

7.4.2. *Other Component Developments*

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

In 2019, several components of CADP have been improved as follows:

- We enhanced the TESTOR tool by adding the possibility to interact with an SUT (*System Under Test*) using its standard input and output.
- We enhanced the XTL compiler with a function converting a transition label into a string (useful for handling the entire content of the label), and we also corrected three bugs.
- We enhanced MCL_EXPAND 5 with a better detection of nondeterminism in probabilistic formulas and a vacuity check for infinite looping operators, and we also corrected a semantic bug.
- We enhanced EVALUATOR 5 with more explanative messages about the assignment of probabilities to transitions, and we corrected two bugs in each of the tools EVALUATOR 4 and 5.
- The C code generated by CAESAR has been modified to suppress GCC 6.5 warnings.
- Several changes have been brought to CADP to enable its use on new platforms, including macOS 10.15 "Catalina" and the forthcoming Debian 10.0 Linux distribution. Various bugs specific to Linux and SunOS systems (Solaris or Illumos/OpenIndiana) have been fixed.

7.5. Real-Life Applications and Case Studies

7.5.1. *Autonomous Resilience of Distributed IoT Applications in a Fog Environment*

Participants: Umar Ozeer, Gwen Salaün.

Recent computing trends have been advocating for more distributed paradigms, namely Fog computing, which extends the capacities of the Cloud at the edge of the network, that is close to end devices and end users in the physical world. The Fog is a key enabler of the Internet of Things (IoT) applications as it resolves some of the needs that the Cloud fails to provide such as low network latencies, privacy, QoS, and geographical requirements. For this reason, the Fog has become increasingly popular and finds application in many fields such as smart homes and cities, agriculture, healthcare, transportation, etc.

The Fog, however, is unstable because it is constituted of billions of heterogeneous devices in a dynamic ecosystem. IoT devices may regularly fail because of bulk production and cheap design. Moreover, the Fog-IoT ecosystem is cyber-physical and thus devices are subjected to external physical world conditions, which increase the occurrence of failures. When failures occur in such an ecosystem, the resulting inconsistencies in the application affect the physical world by inducing hazardous and costly situations.

In the framework of the collaboration with Orange Labs (see § 8.1.1), we proposed an end-to-end autonomic failure management approach for IoT applications deployed in the Fog. The proposed approach recovers from failures in a cyber-physical consistent way. Cyber-physical consistency aims at maintaining a consistent behavior of the application with respect to the physical world, as well as avoiding dangerous and costly circumstances. The approach was validated using model checking techniques to verify important correctness properties. It was then implemented as a framework called F3ARIoT. This framework was evaluated on a smart home application. The results showed the feasibility of deploying F3ARIoT on real Fog-IoT applications as well as its good performances in regards to end user experience.

These results have been published in U. Ozeer's PhD thesis [10] and at an international conference [26]. Another paper was submitted to an international journal.

7.5.2. *Verified Composition and Deployment of IoT Applications*

Participants: Alejandro Martinez Rivero, Radu Mateescu, Ajay Muroor Nadumane, Gwen Salaün.

The Internet of Things (IoT) applications are built by interconnecting everyday objects over internet. As IoT is becoming popular among consumers, the challenge of making IoT applications easy to design and deploy is more relevant than ever. In 2019, we considered this challenge along two perspectives.

- In the framework of the collaboration with Nokia Bell Labs (see § 8.1.2), we focused on helping consumers to easily design IoT applications that are correct, and also support the deployment of these applications. The correctness of the applications is ensured through formal methods and verification techniques.

  Using W3C Web of Things (WoT) specification as the basis of our work, we extended the specification of objects in WoT with a behavioural model. This allows us to describe formally the composition of objects and thus, to verify their behavioural correctness. Typically, an IoT application is defined using Event-Condition-Action (ECA) rules of the type "IF event THEN action". Our work supports users to specify not only the ECA rules, but also the composition of rules using a simple, yet expressive language. This makes possible the construction of advanced compositions, which would have been hard or sometimes impossible to build using simple ECA rules. Finally, users are provided with an easy-to-deploy solution for these advanced compositions. All these steps were implemented and packaged in a tool named MozART, built on top of Mozilla WebThings platform. LNT is used as the formal specification language, and various tools of CADP are used for verifying the composition. Also, an execution engine based on Mozilla WebThings API was built to support the deployment of advanced compositions. The work has led to the preparation of two conference articles.

- Building IoT applications of added-value from a set of available devices with minimal human intervention is one of the main challenges facing the IoT. This is a difficult task that requires models for specifying objects, in addition to user-friendly and reliable composition techniques which in turn prevent the design of erroneous applications.

  In collaboration with Francisco Durán (University of Málaga, Spain), we tackled this problem by first describing IoT applications using abstract models obtained from existing models of concrete devices. Then, we proposed automated techniques for building compositions of devices using a repository of available devices, and an abstract goal of what the user expects from such compositions. Since the number of possible solutions can be quite high, we used both filtering and ranking techniques to provide the most relevant solutions to users. The provided solutions satisfy the given goal and may be analysed with respect to properties such as deadlock-freeness or unmatched send messages. Finally, the application can be deployed using existing execution engines. This work led to a publication in an international conference [20].

7.5.3. Autonomous Car

Participants: Philippe Ledent, Lina Marsso, Radu Mateescu, Wendelin Serwe.

Autonomous vehicles are complex cyber-physical systems that must satisfy critical correctness requirements to increase the safety of road traffic. The validation of autonomous driving is a challenging field because of the complexity of its key components (perception of the environment, scene interpretation, decision making and undertaking of actions) and the intertwinning of physical and software components. In 2019, we considered this challenge along two lines of work.

- From the embedded software perspective, autonomous cars can be considered as GALS systems, which integrate reactive synchronous components that interact asynchronously. The complexity induced by combining synchronous and asynchronous aspects makes GALS systems difficult to develop and debug.

  In the framework of the ARC6 collaboration (see § 9.1.1), we proposed a testing methodology for GALS systems that leverages conformance test generation for asynchronous systems to automatically derive realistic scenarios (inputs constraints and oracles), which are necessary ingredients for the unit testing of individual synchronous components, and are difficult and error-prone to design manually. The methodology consists of several steps (derivation of asynchronous test cases from a GALS model and a test purpose, projection of the complete test graph on a synchronous component, extraction and execution of test scenarios) and was illustrated on a simple, but relevant example inspired by autonomous cars. These results were published in L. Marsso's PhD thesis [9] and at an international conference [25].

- In collaboration with Christian Laugier, Anshul Paigwar, and Alessandro Renzaglia (CHROMA project-team), we proposed a new approach where formal verification is employed to validate systems with probabilistic predictions. We focused on the risk assessment generated by CMCDOT (*Conditional Monte Carlo Dense Occupancy Tracker*), a probabilistic perception system for autonomous cars. CMCDOT provides an environment representation through Bayesian probabilistic occupancy grids and estimates Time-to-Collision probabilities for every static and dynamic part of the grid in the near future. To validate the probabilistic collision risk estimation, we used the CARLA simulator to generate a large number of realistic intersection crossing scenarios with two vehicles. The set of scenarios is then validated using the XTL model checker, by defining appropriate KPIs (*Key Performance Indicators*) as temporal logic formulas and also performing a quantitative analysis. This work led to a publication at an international conference [24].

DEDUCTEAM Project-Team

6. New Results

6.1. Implementation of Dedukti

During his master internship with Frédéric Blanqui and Bruno Barras, Gabriel Hondet developed a new rewrite engine for Dedukti [22]. The algorithm used in the new rewriting engine is formalised and a correctness proof is provided. This algorithm is based on the pattern matching algorithm by Maranget and used in OCaml. It is extended to rewriting rules, λ terms and non linear patterns. Some interesting implementation details are evinced and then we compare the efficiency of the new engine to a naive matching algorithm and to the rewriting engine of Dedukti. The results show that our implementation handles large rewrite systems better that the naive algorithm, and is always better than Dedukti's.

During her internship with Frédéric Blanqui and Emilio Gallego, Houda Mouzoun developed a Dedukti plugin for the VSCode editor.

During his internship with Frédéric Blanqui and Valentin Blot, Jui-Hsuan Wu implemented a prototype algorithm for deciding whether a function defined by rewriting rules is injective or not [23], and also a new algorithm proposed by Frédéric Blanqui for checking that user-defined rewrite rules preserve typing.

Bruno Barras has developed a reduction machine implementing a strong call-by-need strategy for β-reduction and pattern-matching. Higher-order pattern-matching is not yet fully implemented. Regarding efficiency, an exponential speed-up can be observed compared to the current call-by-name implementation on a large class of examples, but a constant slow-down shows on examples where call-by-name is the optimal strategy. With Beniamino Accattoli, he has started studying the correctness of this machine, without pattern-matching. They proved that the machine correctly implements β-reduction, but have no result yet regarding the strategy or the time complexity.

6.2. Theory of $\lambda\Pi$-calculus modulo rewriting

Dependency pairs are a key concept at the core of modern automated termination provers for first-order term rewriting systems. In [14][15], Frédéric Blanqui, Guillaume Genestier and Olivier Hermant introduced an extension of this technique for a large class of dependently-typed higher-order rewriting systems. This improves previous results by Wahlstedt on the one hand and the first author on the other hand to strong normalization and non-orthogonal rewriting systems. This new result has been implemented in the termination-checker SizeChangeTool [17], which participated in the Termination Competition and is used by Dedukti.

During his internship with Frédéric Blanqui and Valentin Blot, Jui-Hsuan Wu designed an algorithm for deciding whether a function defined by rewriting rules is injective or not [23]. This allows to improve the unification algorithm used in Dedukti for inferring types and missing arguments.

The expressiveness of dependent type theory can be extended by identifying types modulo some additional computation rules. But, for preserving the decidability of type-checking or the logical consistency of the system, one must make sure that those user-defined rewriting rules preserve typing. Frédéric Blanqui has developed a new method to check that property using Knuth-Bendix completion. A prototype implementation by Jui-Hsuan Wu is available in Dedukti.

Confluence is a crucial property of rewriting. Gaspard Férey and Jean-Pierre Jouannaud formalized the higher-order rewriting relation on untyped terms implemented in Dedukti and studied various criteria to obtain confluence of higher-order rewrite systems considered together with beta. In particular Von Oostrom's decreasing diagrams technique is applied to multi-steps extensions of simple term rewriting to achieve confluence criteria based on the decidable computation of (orthogonal) higher-order critical pairs. This work assumes left-linearity of rules for now but current work aims at extending these techniques to prove confluence of non-left-linear rule restricted to subsets of terms [20].

Fran cois Thiré has worked on a criterion that would help proving metatheoretical results on Cumulative Type Systems, such as expansion postponment and the equivalence between typed and untyped presentations of conversion. This has been published andpresented at LFMTP'19 [19]

Frédéric Gilbert has written a preprint about the definition of proof certificates for predicative subtyping [21].

6.3. Proof reconstruction

Proof assistants often call automated theorem provers to prove subgoals. However, each prover has its own proof calculus and the proof traces that it produces often lack many details to build a complete proof. Hence these traces are hard to check and reuse in proof assistants. Dedukti is a proof checker whose proofs can be translated to various proof assistants: Coq, HOL, Lean, Matita, PVS. Yacine El Haddad, Guillaume Burel and Frédéric Blanqui implemented a tool Ektraskto [16] that extracts TPTP subproblems from a TSTP file and reconstructs complete proofs in Dedukti using automated provers able to generate Dedukti proofs like ZenonModulo or ArchSAT. This tool is generic: it assumes nothing about the proof calculus of the prover producing the trace, and it can use different provers to produce the Dedukti proof. We applied our tool on traces produced by automated theorem provers on the CNF problems of the TPTP library and we were able to reconstruct a proof for a large proportion of them, significantly increasing the number of Dedukti proofs that could be obtained for those problems.

Zenon Modulo and iProverModulo, two automated theorem provers that can produce Dedukti proofs, have been presented in an article accepted in the Journal of Automated Reasoning [12].

6.4. Translating proofs to Dedukti

Agda is a dependently-typed programming language developed at Chalmers University, Gothenburg, Sweden, for 20 years. Thanks to the propositions-as-types correspondence of Curry-Howard, Agda is often used as a proof-assistant. Guillaume Genestier developed with Jesper Cockx a prototypical translator from Agda to Dedukti, which supports well some of the mainly used features of Agda and translates hundreds of definitions of the standard libraries. This implementation led to new encodings of theories in Dedukti, regarding: Universe Polymorphism, Inductive and Record Types, Dependent Pattern Matching, eta convertibility. The implementation of this translator permits to improve both Agda and Dedukti. Indeed, we discovered some bugged (almost not used) functions in Agda and had to extend some existing functions to our purpose. On the Dedukti side, this implementation was the first usage of the newly implemented feature of rewriting modulo associativity and commutativity, which required some minor improvements. Furthermore, our translation of eta-expansion using a defined function led to an improvement in the verification of type preservation of rewriting rules in Dedukti.

Isabelle is a logical framework developed at Technical University of Munich and Cambridge University since the 90s. It implements several logics such as HOL and ZF and is used as part of large verification projects such as seL4 and Flyspeck. Gabriel Hondet developed with Makarius Wenzel (from Augsburg) an export from Isabelle propositions to Dedukti, which was later extended by Michael Färber and Makarius Wenzel to export proofs. This required substantial work on the Isabelle kernel to extend the reconstruction of proof terms based on the work of Stefan Berghofer. Our newly developed proof export allows for an independent verification of a substantial portion of the Isabelle/HOL standard library as well as for the integration of results proved in Isabelle into Logipedia.

6.5. Models of cubical type theory

Bruno Barras and Rehan Malak have developed further their Dedukti library of presheaves. Using this library, they have built a semi-simplicial model of System F.

6.6. A proof system for PCTL and CTL*

Gilles Dowek, Ying Jiang, and Wu Peng, have proposed a proof system for the probabilistic modal logic PCTL. A paper is in preparation.

Gilles Dowek, Ying Jiang, Wu Peng, and Wenhui Zhang have started to study a proof system for CTL*, that mixes constructive and classical aspects.

The article Towards Combining Model Checking and Proof Checking, of Ying Jiang, Jian Liu, Gilles Dowek, and Kailiang Ji, has been published in The Computer Journal [13].

6.7. System I

Gilles Dowek and Alejandro Díaz-Caro have defined a lambda-calculus, the system I, to represent the proofs of a variant minimal propositional logic where isomorphic propositions are identified. Their paper Proof Normalisation in a Logic Identifying Isomorphic Propositions, has been presented at the International Conference on Formal Structures for Computation and Deduction. A second paper The virtues of eta-expansion in System I, showing that the addition of eta-expansion to system I actually simplies the system has been submitted to publication.

6.8. Computing with global environments

The call-by-need evaluation strategy for the λ-calculus is an evaluation strategy that lazily evaluates arguments only if needed, and if so, shares computations across all places where it is needed. To implement this evaluation strategy, abstract machines require some form of global environment. While abstract machines usually lead to a better understanding of the flow of control during the execution, easing in particular the definition of continuation-passing style translations, the case of machines with global environments turns out to be much more subtle.

In collaboration with Hugo Herbelin, Étienne Miquey introduced F_{Υ}, a calculus featuring a data type for typed stores and a mechanism of explicit coercions witnessing store extensions. This calculus defines a generic target of typed continuationand-environment-passing style translations for several calculi with global environment: it is compatible with different evaluation strategy (call-by-need, call-by-name, call-by-value) and different type systems (simple types, system F). On the logical side, these translations broadly amounts to a Kripke forcing-like translation mixed with a negative translation (for the continuation-passing part).

6.9. Computional interpretation of the axiom scheme of comprehension

The axiom scheme of comprehension is the cornerstone of second-order arithmetic, a logical theory in which most of mathematics can be formalized. Historically, comprehension was obtained from the negative translation of the axiom of choice, this axiom being interpreted by bar recursion. This led to cluttered and inefficient interpretations of second-order arithmetic.

Valentin Blot simplified this interpretation by proving that the axiom scheme of comprehension has a direct computational interpretation through a variant of bar recursion called update recursion. This new interpretation leads to a more efficient computational interpretation of proofs in second-order arithmetic, and paves the way for a convergence of the two existing interpretations: bar recursion and System F.

6.10. Alignment of logical connectives

Émilie Grienenberger and Gilles Dowek have studied in practice the alignment of logical connectives between proofs systems, a first step towards concept alignement, by the export of the HOL Light standard library using axiomatized connectives to Dedukti. More theoretically, an ecumenical system–where classical and intuitionistic logics coexist–was introduced to act as an exchange platform between proof systems.

6.11. Quantum Computing

The article Two linearities for quantum computing in the lambda calculus, of Alejandro Díaz-Caro, Gilles Dowek, and Juan Pablo Rinaldi, first published in the proceedings of Theory and Practice of Natural Computing 2017, has been published in the journal Biosystems.

GALLINETTE Project-Team

6. New Results

6.1. Logical Foundations of Programming Languages

Participants: Esaie Bauer, Rémi Douence, Marie Kerjean, Ambroise Lafont, Maxime Lucas, Étienne Miquey, Guillaume Munch-Maccagnoni, Nicolas Tabareau.

6.1.1. Classical Logic

6.1.1.1. Continuation-and-environment-passing style translations: a focus on call-by-need

The call-by-need evaluation strategy for the λ-calculus is an evaluation strategy that lazily evaluates arguments only if needed, and if so, shares computations across all places where it is needed. To implement this evaluation strategy, abstract machines require some form of global environment. While abstract machines usually lead to a better understanding of the flow of control during the execution, easing in particular the definition of continuation-passing style translations, the case of machines with global environments turns out to be much more subtle. The main purpose of [21] is to understand how to type a continuation-and-environment-passing style translations, that it to say how to soundly translate a classical calculus with environment into a calculus that does not have these features. To this end, we focus on a sequent calculus presentation of a call-by-need λ-calculus with classical control for which Ariola et. al already defined an untyped translation and which we equipped with a system of simple types in a previous paper. We present here a type system for the target language of their translation, which highlights a variant of Kripke forcing related to the environment-passing part of the translation. Finally, we show that our construction naturally handles the cases of call-by-name and call-by-value calculi with environment, encompassing in particular the Milner Abstract Machine, a machine with global environments for the call-by-name λ-calculus.

6.1.1.2. Revisiting the duality of computation: an algebraic analysis of classical realizability models

In an impressive series of papers, Krivine showed at the edge of the last decade how classical realizability provides a surprising technique to build models for classical theories. In particular, he proved that classical realizability subsumes Cohen's forcing, and even more, gives rise to unexpected models of set theories. Pursuing the algebraic analysis of these models that was first undertaken by Streicher, Miquel recently proposed to lay the algebraic foundation of classical realizability and forcing within new structures which he called implicative algebras. These structures are a generalization of Boolean algebras based on an internal law representing the implication. Notably, implicative algebras allow for the adequate interpretation of both programs (i.e. proofs) and their types (i.e. formulas) in the same structure. The very definition of implicative algebras takes position on a presentation of logic through universal quantification and the implication and, computationally, relies on the call-by-name λ-calculus. In [13], we investigate the relevance of this choice, by introducing two similar structures. On the one hand, we define disjunctive algebras, which rely on internal laws for the negation and the disjunction and which we show to be particular cases of implicative algebras. On the other hand, we introduce conjunctive algebras, which rather put the focus on conjunctions and on the call-by-value evaluation strategy. We finally show how disjunctive and conjunctive algebras algebraically reflect the well-known duality of computation between call-by-name and call-by-value.

6.1.2. Models of programming languages mixing effects and resources

6.1.2.1. Efficient deconstruction with typed pointer reversal

Building on the connection between resource management in systems programming and ordered logic we established previously, we investigate a pervasive issue in the languages C++ and Rust whereby compiler-generated clean-up functions cause a stack overflow on deep structures. In [17], we show how to generate clean-up algorithms that run in constant time and space for a broad class of ordered algebraic datatypes such as ones that can be found in C++ and Rust or in future extensions of functional programming languages with first-class resources.

6.1.2.2. Resource safety in OCaml

Building on our investigations for a resource-management model for OCaml, we have proposed several preliminary improvements to the OCaml language. We contributed to the design and implementation of new resource management primitives (PRs #2118, #8962), resource-safe C APIs (PRs #8993, #8997, #9037), and core runtime capabilities (PR #8961). (#2118 has been merged into OCaml 4.08 and #8993 and #9037 have been merged into OCaml 4.10.)

We continued to interact with L. White and S. Dolan (Jane Street), on the design of resource management and exception safety in multicore OCaml.

6.1.3. Syntax and Rewriting Systems

6.1.3.1. Reduction Monads and Their Signatures

In [1], we study reduction monads, which are essentially the same as monads relative to the free functor from sets into multigraphs. Reduction monads account for two aspects of the lambda calculus: on the one hand, in the monadic viewpoint, the lambda calculus is an object equipped with a well-behaved substitution; on the other hand, in the graphical viewpoint, it is an oriented multigraph whose vertices are terms and whose edges witness the reductions between two terms. We study presentations of reduction monads. To this end, we propose a notion of reduction signature. As usual, such a signature plays the role of a virtual presentation, and specifies arities for generating operations-possibly subject to equations-together with arities for generating reduction rules. For each such signature, we define a category of models; any model is, in particular, a reduction monad. If the initial object of this category of models exists, we call it the reduction monad presented (or specified) by the given reduction signature. Our main result identifies a class of reduction signatures which specify a reduction monad in the above sense. We show in the examples that our approach covers several standard variants of the lambda calculus.

6.1.3.2. Modules over monads and operational semantics

[22] is a contribution to the search for efficient and high-level mathematical tools to specify and reason about (abstract) programming languages or calculi. Generalising the reduction monads of Ahrens et al., we introduce operational monads, thus covering new applications such as the-calculus, Positive GSOS specifications, and the big-step, simply-typed, call-by-value-calculus. Finally, we design a notion of signature for operational monads that covers all our examples.

6.1.3.3. Modular specification of monads through higher-order presentations

In their work on second-order equational logic, Fiore and Hur have studied presentations of simply typed languages by generating binding constructions and equations among them. To each pair consisting of a binding signature and a set of equations, they associate a category of 'models', and they give a monadicity result which implies that this category has an initial object, which is the language presented by the pair. In [10], we propose, for the untyped setting, a variant of their approach where monads and modules over them are the central notions. More precisely, we study, for monads over sets, presentations by generating ('higher-order') operations and equations among them. We consider a notion of 2-signature which allows to specify a monad with a family of binding operations subject to a family of equations, as is the case for the paradigmatic example of the lambda calculus, specified by its two standard constructions (application and abstraction) subject to β- and η-equalities. Such a 2-signature is hence a pair (Σ,E) of a binding signature Σ and a family E of equations for Σ. This notion of 2-signature has been introduced earlier by Ahrens in a slightly different context. We associate, to each 2-signature (Σ,E), a category of 'models of (Σ,E); and we say that a 2-signature is 'effective' if this category has an initial object; the monad underlying this (essentially unique) object is the 'monad specified by the 2-signature'. Not every 2-signature is effective; we identify a class of 2-signatures, which we call 'algebraic', that are effective. Importantly, our 2-signatures together with their models enjoy 'modularity': when we glue (algebraic) 2-signatures together, their initial models are glued accordingly. We provide a computer formalization for our main results.

6.1.3.4. The Diamond Lemma for non-terminating rewriting systems

In [16], we study the confluence property for rewriting systems whose underlying set of terms admits a vector space structure. For that, we use deterministic reduction strategies. These strategies are based on the choice of standard reductions applied to basis elements. We provide a sufficient condition of confluence in terms of the kernel of the operator which computes standard normal forms. We present a local criterion which enables us to check the confluence property in this framework. We show how this criterion is related to the Diamond Lemma for terminating rewriting systems

6.1.4. Differential Linear Logic

6.1.4.1. Higher-order distributions for differential linear logic

Linear Logic was introduced as the computational counterpart of the algebraic notion of linearity. Differential Linear Logic refines Linear Logic with a proof-theoretical interpretation of the geometrical process of differentiation. In [24], we construct a polarized model of Differential Linear Logic satisfying computational constraints such as an interpretation for higher-order functions, as well as constraints inherited from physics such as a continuous interpretation for spaces. This extends what was done previously by Kerjean for first order Differential Linear Logic without promotion. Concretely, we follow the previous idea of interpreting the exponential of Differential Linear Logic as a space of higher-order distributions with compact-support, and is constructed as an inductive limit of spaces of distributions on Euclidean spaces. We prove that this exponential is endowed with a co-monadic like structure, with the notable exception that it is functorial only on isomorphisms. Interestingly, as previously argued by Ehrhard, this still allows one to interpret differential linear logic without promotion.

6.1.4.2. Chiralities in topological vector spaces

Chiralities are categories introduced by Mellies to account for a game semantics point of view on negation. In [23], [20], we uncover instances of this structure in the theory of topological vector spaces, thus constructing several new polarized models of Multiplicative Linear Logic. These models improve previously known smooth models of Differential Linear Logic, showing the relevance of chiralities to express topological properties of vector spaces. They are the first denotational polarized models of Multiplicative Linear Logic, based on the pre-existing theory of topological vector spaces, in which two distinct sets of formulas, two distinct negations, and two shifts appear naturally.

6.1.5. Distributed Programming

6.1.5.1. 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 [9], 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, Gaëtan Gilbert, Maxime Lucas, Pierre-Marie Pédrot, Loïc Pujet, Nicolas Tabareau, Théo Winterhalter.

6.2.1. Type Theory

6.2.1.1. Effects in Type Theory.

There is a critical tension between substitution, dependent elimination and effects in type theory. In this paper, we crystallize this tension in the form of a no-go theorem that constitutes the fire triangle of type theory. To release this tension, we propose in [7] DCBPV, an extension of call-by-push-value (CBPV)-a general calculus of effects-to dependent types. Then, by extending to CBPV the well-known decompositions of call-by-name and call-by-value into CBPV, we show why, in presence of effects, dependent elimination must be restricted in call-by-name, and substitution must be restricted in call-by-value. To justify DCBPV and show that it is general enough to interpret many kinds of effects, we define various effectful syntactic translations from DCBPV to Martin-Löf type theory: the reader, weaning and forcing translations.

Traditional approaches to compensate for the lack of exceptions in type theories for proof assistants have severe drawbacks from both a programming and a reasoning perspective. We recently extended the Calculus of Inductive Constructions (CIC) with exceptions. The new exceptional type theory is interpreted by a translation into CIC, covering full dependent elimination, decidable type-checking and canonicity. However, the exceptional theory is inconsistent as a logical system. To recover consistency, we propose an additional translation that uses parametricity to enforce that all exceptions are caught locally. While this enforcement brings logical expressivity gains over CIC, it completely prevents reasoning about exceptional programs such as partial functions. In [6], we addresses the dilemma between exceptions and consistency in a more flexible manner, with the Reasonably Exceptional Type Theory (RETT). RETT is structured in three layers: (a) the exceptional layer, in which all terms can raise exceptions; (b) the mediation layer, in which exceptional terms must be provably parametric; (c) the pure layer, in which terms are non-exceptional, but can refer to exceptional terms. We present the general theory of RETT, where each layer is realized by a predicative hierarchy of universes, and develop an instance of RETT in Coq: the impure layer corresponds to the predicative universe hierarchy, the pure layer is realized by the impredicative universe of propositions, and the mediation layer is reified via a parametricity type class. RETT is the first full dependent type theory to support consistent reasoning about exceptional terms, and the CoqRETT plugin readily brings this ability to Coq programmers.

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 [15], 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. Setoid type theory - a syntactic translation

[11] introduces setoid type theory, an intensional type theory with a proof-irrelevant universe of propositions and an equality type satisfying function extensionality, propositional extensionality and a definitional computation rule for transport. We justify the rules of setoid type theory by a syntactic translation into a pure type theory with a universe of propositions. We conjecture that our syntax is complete with regards to this translation.

6.2.1.4. The folk model category structure on strict ω-categories is monoidal

In [19], we prove that the folk model category structure on the category of strict ω-categories, introduced by Lafont, Métayer and Worytkiewicz, is monoidal, first, for the Gray tensor product and, second, for the join of ω-categories, introduced by the first author and Maltsiniotis. We moreover show that the Gray tensor product induces, by adjunction, a tensor product of strict (m, n)-categories and that this tensor product is also compatible with the folk model category structure. In particular, we get a monoidal model category structure on the category of strict ω-groupoids. We prove that this monoidal model category structure satisfies the monoid axiom, so that the category of Gray monoids, studied by the second author, bears a natural model category structure.

6.2.2. Proof Assistants

6.2.2.1. Metacoq

The MetaCoq project [26], [26] aims to provide a certified meta-programming environment in Coq. It builds on Template-Coq, a plugin for Coq originally implemented by Malecha (2014), which provided a reifier for Coq terms and global declarations, as represented in the Coq kernel, as well as a denotation command. Recently, it was used in the CertiCoq certified compiler project (Anand et al., 2017), as its front-end language, to derive parametricity properties (Anand and Morrisett, 2018). 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 generalize it to handle the entire Polymorphic Calculus of Cumulative Inductive Constructions (pCUIC), 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 and a certifying extraction to call-by-value λ-calculus. We also advocate the use of MetaCoq as a foundation for higher-level tools.

6.2.2.2. Verification of Type Checking and Erasure for Coq, in Coq

Coq is built around a well-delimited kernel that perfoms typechecking for definitions in a variant of the Calculus of Inductive Constructions (CIC). Although the metatheory of CIC is very stable and reliable, the correctness of its implementation in Coq is less clear. Indeed, implementing an efficient type checker for CIC is a rather complex task, and many parts of the code rely on implicit invariants which can easily be broken by further evolution of the code. Therefore, on average, one critical bug has been found every year in Coq. [8] presents the first implementation of a type checker for the kernel of Coq (without the module system and template polymorphism), which is proven correct in Coq with respect to its formal specification and axiomatisation of part of its metatheory. Note that because of Gödel's incompleteness theorem, there is no hope to prove completely the correctness of the specification of Coq inside Coq (in particular strong normalisation or canonicity), but it is possible to prove the correctness of the implementation assuming the correctness of the specification, thus moving from a trusted code base (TCB) to a trusted theory base (TTB) paradigm. Our work is based on the MetaCoq project which provides metaprogramming facilities to work with terms and declarations at the level of this kernel. Our type checker is based on the specification of the typing relation of the Polymorphic, Cumulative Calculus of Inductive Constructions (pCUIC) at the basis of Coq and the verification of a relatively efficient and sound type-checker for it. In addition to the kernel implementation, an essential feature of Coq is the so-called extraction: the production of executable code in functional languages from Coq definitions. We present a verified version of this subtle type-and-proof erasure step, therefore enabling the verified extraction of a safe type-checker for Coq.

6.2.2.3. 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.

6.2.2.4. Cubical Synthetic Homotopy Theory

Homotopy type theory is an extension of type theory that enables synthetic reasoning about spaces and homotopy theory. This has led to elegant computer formalizations of multiple classical results from homotopy theory. However, many proofs are still surprisingly complicated to formalize. One reason for this is the axiomatic treatment of univalence and higher inductive types which complicates synthetic reasoning as many intermediate steps, that could hold simply by computation, require explicit arguments. Cubical type theory offers a solution to this in the form of a new type theory with native support for both univalence and higher inductive types. In [14], we show how the recent cubical extension of Agda can be used to formalize some of the major results of homotopy type theory in a direct and elegant manner.

6.3. Program Certifications and Formalisation of Mathematics

Participants: Julien Cohen, Rémi Douence, Guilhem Jaber, Assia Mahboubi, Igor Zhirkov.

6.3.1. CoqTL: A Coq DSL for Rule-Based Model Transformation

In model-driven engineering, model transformation (MT) verification is essential for reliably producing software artifacts. While recent advancements have enabled automatic Hoare-style verification for non-trivial MTs, there are certain verification tasks (e.g. induction) that are intrinsically difficult to automate. Existing tools that aim at simplifying the interactive verification of MTs typically translate the MT specification (e.g. in ATL) and properties to prove (e.g. in OCL) into an interactive theorem prover. However, since the MT specification and proof phases happen in separate languages, the proof developer needs a detailed knowledge of the translation logic. Naturally, any error in the MT translation could cause unsound verification, i.e. the MT executed in the original environment may have different semantics from the verified MT. In [2] , we propose an alternative solution by designing and implementing an internal domain specific language, namely CoqTL, for the specification of declarative MTs directly in the Coq interactive theorem prover. Expressions in CoqTL are written in Gallina (the specification language of Coq), increasing the possibilities of reusing native Coq libraries in the transformation definition and proof. CoqTL specifications can be directly executed by our transformation engine encoded in Coq, or a certified implementation of the transformation can be generated by the native Coq extraction mechanism. We ensure that CoqTL has the same expressive power of Gallina (i.e. if a MT can be computed in Gallina, then it can also be represented in CoqTL). In this article, we introduce CoqTL, evaluate its practical applicability on a use case, and identify its current limitations.

6.3.2. A certificate-based approach to formally verified approximations.

In [12], we present a library to verify rigorous approximations of univariate functions on real numbers, with the Coq proof assistant. Based on interval arithmetic, this library also implements a technique of validation a posteriori based on the Banach fixed-point theorem. We illustrate this technique on the case of operations of division and square root. This library features a collection of abstract structures that organise the specification of rigorous approximations, and modularise the related proofs. Finally, we provide an implementation of verified Chebyshev approximations, and we discuss a few examples of computations.

6.3.3. 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.

6.3.4. Reasoning about exact memory transformations induced by refactorings in CompCert C

[18] reports on our work in extending CompCert memory model with a relation to model relocations. It preserves undefined values unlike similar relations defined in CompCert. This relation commutes with memory operations. Our main contributions are the relation itself and mechanically checked proofs of its commutation properties. We intend to use this extension to construct and verify a refactoring tool for programs written in C.

6.3.5. Automating Contextual Equivalence for Higher-Order Programs with References

In [4], we have proposed a framework to study contextual equivalence of programs written in a call-by-value functional language with local integer references. It reduces the problem of contextual equivalence to the problem of non-reachability in a transition system of memory configurations. This reduction is complete for recursion-free programs. Restricting to programs that do not allocate references inside the body of functions, we have encoded this non-reachability problem as a set of constrained Horn clause that can then be checked for satisfiability automatically. Restricting furthermore to a language with finite data-types, we also get a new decidability result for contextual equivalence at any type.

MEXICO Project-Team

7. New Results

7.1. Generalized Alignment-Based Trace Clustering of Process Behavior

Process mining techniques use event logs containing real process executions in order to mine, align and extend process models. The partition of an event log into trace variants facilitates the understanding and analysis of traces, so it is a common pre-processing in process mining environments. Trace clustering automates this partition; traditionally it has been applied without taking into consideration the availability of a process model. In this paper we extend our previous work on process model based trace clustering, by allowing cluster centroids to have a complex structure, that can range from a partial order, down to a subnet of the initial process model. This way, the new clustering framework presented in [28] is able to cluster together traces that are distant only due to concurrency or loop constructs in process models. We show the complexity analysis of the different instantiations of the trace clustering framework, and have implemented it in a prototype tool that has been tested on different datasets.

7.2. The involution tool for accurate digital timing and power analysis

In [23] we introduce the prototype of a digital timing simulation and power analysis tool for integrated circuit (Involution Tool) which employs the involution delay model introduced by Fuegger et al. at DATE'15. Unlike the pure and inertial delay models typically used in digital timing analysis tools, the involution model faithfully captures pulse propagation. The presented tool is able to quantify for the first time the accuracy of the latter by facilitating comparisons of its timing and power predictions with both SPICE-generated results and results achieved by standard timing analysis tools. It is easily customizable, both w.r.t. different instances of the involution model and different circuits, and supports automatic test case generation, including parameter sweeping. We demonstrate its capabilities by providing timing and power analysis results for three circuits in varying technologies: an inverter tree, the clock tree of an open-source processor, and a combinational circuit that involves multi-input NAND gates. It turns out that the timing and power predictions of two natural types of involution models are significantly better than the predictions obtained by standard digital simulations for the inverter tree and the clock tree. For the NAND circuit, the performance is comparable but not significantly better. Our simulations thus confirm the benefits of the involution model, but also demonstrate shortcomings for multi-input gates.

7.3. Transistor-level analysis of dynamic delay models

Delay estimation is a crucial task in digital circuit design as it provides the possibility to assure the desired func-tionality, but also prevents undesired behavior very early. For this purpose elaborate delay models like the Degradation Delay Model (DDM) and the Involution Delay Model (IDM) have been proposed in the past, which facilitate accurate dynamic timing analysis: Both use delay functions that determine the delay of the current input transition based on the time difference T to the previous output one. Currently, however, extensive analog simulations are necessary to determine the (parameters of the) delay function, which is a very time-consuming and cumbersome task and thus limits the applicability of these models. In [21], we therefore thoroughly investigate the characterization procedures of a CMOS inverter on the transistor level in order to derive analytical expressions for the delay functions. Based on reasonably simple transistor models we identify three operation regions, each described by a different estimation function. Using simulations with two independent technologies, we show that our predictions are not only accurate but also reasonably robust w.r.t. variations. Our results furthermore indicate that the exponential fitting proposed for DDM is actually only partially valid, while our analytic approach can be applied on the whole range. Even the more complex IDM is predicted reasonably accurate.

7.4. A faithful binary circuit model

[Fuegger et al., IEEE TC 2016] proved that no existing digital circuit model, including those based on pure and inertial delay channels, faithfully captures glitch propagation: For the Short-Pulse Filtration (SPF) problem similar to that of building a one-shot inertial delay, they showed that every member of the broad class of bounded single-history channels either contradicts the unsolvability of SPF in bounded time or the solvability of SPF in unbounded time in physical circuits. In [12], we propose binary circuit models based on novel involution channels that do not suffer from this deficiency. Namely, in sharp contrast to bounded single-history channels, SPF cannot be solved in bounded time with involution channels, whereas it is easy to provide an unbounded SPF implementation. Hence, binary-valued circuit models based on involution channels allow to solve SPF precisely when this is possible in physical circuits. Additionally, using both Spice simulations and physical measurements of an inverter chain instrumented by high-speed analog amplifiers, we demonstrate that our model provides good modeling accuracy with respect to real circuits as well. Consequently , our involution channel model is not only a promising basis for sound formal verification, but also allows to seamlessly improve existing dynamic timing analysis.

7.5. 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 component updates can have a dramatic impact on the predicted behaviours. In [10], we explore the use of Read (contextual) Petri Nets (RPNs) to study dynamics of BNs from a concurrency theory perspective. After showing bi-directional translations between RPNs and BNs and analogies between results on synchronism sensitivity, we illustrate that usual updating modes for BNs can miss plausible behaviours, i.e., incorrectly conclude on the absence/impossibility of reaching specific configurations. We propose an encoding of BNs capitalizing on the RPN semantics enabling more behaviour than the generalized asynchronous updating mode. The proposed encoding ensures a correct abstraction of any multivalued refinement, as one may expect to achieve when modelling biological systems with no assumption on its time features.

7.6. Sequential Reprogramming of Boolean Networks Made Practical

We address the sequential reprogramming of gene regulatory networks modelled as Boolean networks.

- Cellular reprogramming, a technique that opens huge opportunities in modern and regenerative medicine, heavily relies on identifying key genes to perturb. Most of the existing computational methods for controlling which attractor (steady state) the cell will reach focus on finding mutations to apply to the initial state. However, it has been shown, and is proved in our article [14], that waiting between perturbations so that the update dynamics of the system prepares the ground, allows for new reprogramming strategies. To identify such sequential perturbations, we consider a qualitative model of regulatory networks, and rely on Binary Decision Diagrams to model their dynamics and the putative perturbations. Our method establishes a set identification of sequential perturbations, whether permanent (mutations) or only temporary, to achieve the existential or inevitable reachability of an arbitrary state of the system. We apply an implementation for temporary perturbations on models from the literature, illustrating that we are able to derive sequential perturbations to achieve trans-differentiation.
- In [22], we develop an attractor-based sequential reprogramming method to compute all sequential reprogramming paths from a source attractor to a target attractor, where only attractors of the network are used as intermediates. Our method is more practical than existing reprogramming methods as it incorporates several practical constraints: (1) only biologically observable states, viz. attractors, can act as intermediates; (2) certain attractors, such as apoptosis, can be avoided as intermediates; (3) certain nodes can be avoided to perturb as they may be essential for cell survival or difficult to perturb with biomolecular techniques; and (4) given a threshold k, all sequential reprogramming paths with no more than k perturbations are computed. We compare our method with the minimal one-step

reprogramming and the minimal sequential reprogramming on a variety of biological networks. The results show that our method can greatly reduce the number of perturbations compared to the one-step reprogramming, while having comparable results with the minimal sequential reprogramming. Moreover, our implementation is scalable for networks of more than 60 nodes.

7.7. 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. In [13], we introduce 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.

7.8. Combining Refinement of Parametric Models with Goal-Oriented Reduction of Dynamics

Parametric models abstract part of the specification of dynamical models by integral parameters. They are for example used in computational systems biology, notably with parametric regulatory networks, which specify the global architecture (interactions) of the networks, while parameterising the precise rules for drawing the possible temporal evolutions of the states of the components. A key challenge is then to identify the discrete parameters corresponding to concrete models with desired dynamical properties. Our work [20] addresses the restriction of the abstract execution of parametric regulatory (discrete) networks by the means of static analysis of reachability properties (goal states). Initially defined at the level of concrete parameterised models, the goal-oriented reduction of dynamics is lifted to parametric networks, and is proven to preserve all the minimal traces to the specified goal states. It results that one can jointly perform the refinement of parametric networks (restriction of domain of parameters) while reducing the necessary transitions to explore and preserving reachability properties of interest.

7.9. Autonomous Transitions Enhance CSLTA Expressiveness and Conciseness

CSLTA is a stochastic temporal logic for continuous-time Markov chains (CTMC) where formulas similarly to those of CTL* are inductively defined by nesting of timed path formulas and state formulas. In particular a timed path formula of CSLTA is specified by a single-clock Deterministic Timed Automaton (DTA). Such a DTA features two kinds of transitions: synchronizing transitions triggered by CTMC transitions and autonomous transitions triggered by time elapsing that change the location of the DTA when the clock reaches a given threshold. It has already been shown that CSLTA strictly includes stochastic logics like CSL and asCSL. An interesting variant of CSLTA consists in equipping transitions rather than locations by boolean formulas. In [27], we answer the following question: do autonomous transitions and/or boolean guards on transitions enhance expressiveness and/or conciseness of DTAs? We show that this is indeed the case. In establishing our main results we also identify an accurate syntactical characterization of DTAs for which the autonomous transitions do not add expressive power but lead to exponentially more concise DTAs.

7.10. Coverability and Termination in Recursive Petri Nets

In the early two-thousands, Recursive Petri nets have been introduced in order to model distributed planning of multi-agent systems for which counters and recursivity were necessary. Although Recursive Petri nets strictly extend Petri nets and stack automata, most of the usual property problems are solvable but using non primitive recursive algorithms, even for coverability and termination. For almost all other extended Petri nets models containing a stack the complexity of coverability and termination are unknown or strictly larger than EXPSPACE. In contrast, we establish in [18] that for Recursive Petri nets, the coverability and termination problems are EXPSPACE-complete as for Petri nets. From an expressiveness point of view, we show that coverability languages of Recursive Petri nets strictly include the union of coverability languages of Petri nets and context-free languages. Thus we get for free a more powerful model than Petri net.

MOCQUA Team

7. New Results

7.1. Semicomputable points in Euclidean spaces

- Participants: Mathieu Hoyrup, Donald Stull

Many natural problems/objects from theoretical computer science and logic are not decidable/computable, but semidecidable/semicomputable only: the halting problem, provability, domino problem, attractors of dynamical systems, etc. We pursue our program to study semicomputable objects in a systematic way. In this work, we focus on objects that can be described by finitely many real numbers, in particular polynomials and disks in the plane. Such objects can be identified with points of Euclidean spaces. We therefore introduce and study a notion of semicomputable point in Euclidean spaces, providing a multi-dimensional analog of a well-known unidimensional notion. The study involves ideas from linear algebra, convex analysis and computability. This work was presented at MFCS 2019 [27].

7.2. Computability on quasi-Polish spaces

- Participants: Mathieu Hoyrup, Cristobal Rojas, Victor Selivanov, Donald Stull

Descriptive Set Theory (DST) is a branch of topology which interacts very nicely with computability and logic. Indeed, these three theories involve measuring the complexity of describing objects in different ways (respectively as combinations of open sets, by programs, by formulae), which are intimately related. However, DST is traditionally developed on spaces relevant to mathematical analysis (Polish spaces), but not to theoretical computer science. The recently introduced quasi-Polish spaces are a much broader class of spaces including for instance Scott domains, important in functional programming. However, how to compute in such spaces is still not well-understood. In particular, quasi-Polish spaces can be characterized in many ways, so one has to choose the right definition to start with. We compare the computable versions of some of them, proving their non-equivalence, and focus on one of them, providing evidence that this notion is probably the right one. This work was presented at DCFS 2019 [26].

7.3. Degree spectra of Polish spaces

- Participants: Mathieu Hoyrup, Takayuki Kihara, Victor Selivanov

Mathematical objects can encode information. An obvious example is given by subsets of the plane: a text printed on a sheet of paper is a subset of the plane conveying information. However, when the object is submitted to deformations, what information can still be conveyed? What information is invariant under such deformations?

It is the core question in computable structure theory: for instance, what can be encoded in an infinite graph, which can be decoded from the structure itself and not from a particular presentation of the graph? Mathematically, what information is robust under graph isomorphism? It happens that much information can be encoded, for instance by using the lengths of the cycles in the graph.

Albegraic structures have been thoroughly studied from this perspective. However, the study of topological structures is almost inexistant, and more difficult (they are continuous while algebraic structures are often discrete). For instance, what information can be encoded in a subset of the plane, which is stable under continuous deformations (homeomorphisms)?

We have tackled this question during the visit of Takayuki Kihara and Victor Selivanov, and obtained many interesting results. For instance, we have proved that no direct information can be encoded (for instance, no infinite binary sequence can be extracted by an algorithm, unless the sequence is already computable). However, limit information can be encoded (for instance, a binary sequence can be encoded in such a way that a double-sequence converging to it can be extracted from the object by an algorithm). It is still open whether a single limit is possible.

A paper is still in preparation.

7.4. Computable SFTs

- Participants: Emmanuel Jeandel and Pascal Vanier

Previous works by the two participants have shown that there is a striking similarity between subshifts of finite type (tilings, coloring of the plane that do not contain a given set of patterns) and finitely presented groups (finitely generated groups with a finite number of equations).

This analogy can be described intuitively as follows: colors in subshifts corresponds to the generators of the groups, forbidden patterns correspond to the equations. Finite type is the same as finite presentation, and minimal subshifts correspond to simple groups.

The article [29] develops this analogy to computable objects: It is well known by the Higman-Thompson theorem that a finitely generated group is computable iff it is a subgroup of a simple group which is itself a subgroup of a finitely presented group. In this article, we give an equivalent for subshifts : a subshift is computable iff it is the restriction of a minimal subshift which is itself the restriction of a subshift of finie type.

7.5. Probabilistic cellular automata for problem solving

- Participants: Nazim Fatès, Irène Marcovici

Directly related to the theme exposed in Sec. 4.3 , we examined the problem of self-stabilisation, as introduced by Dijkstra in the 1970's, in the context of cellular automata [33]. More precisely, we examined how to stabilise k-colourings, that is, infinite grids which are coloured with k distinct colours in such a way that adjacent cells have different colours. The idea is that if, for any reason (e.g., noise, previous usage, tampering by an adversary), the colours of a finite number of cells in a valid k-colouring are modified, thus introducing errors, we can correct the system into a valid k-colouring by using local rules only. In other words, we designed cellular automaton rules which, starting from any finite perturbation of a valid k-colouring, reach a valid k-colouring in finite time. We discussed the different cases depending on the number of colours k, and propose some deterministic and probabilistic rules which solve the problem for $k \neq 3$. We also explained why the case $k = 3$ is more delicate. Finally, we proposed some insights on the more general setting of this problem, passing from k-colourings to other tilings (subshifts of finite type).

In the same spirit, we addressed the problem of detecting failures in a distributed network [30]. Our question is: if some components can break down over time, how can we detect that the failure rate has exceeded a given threshold without any central authority? We want to estimate the global state of the network, only through local interactions of components with their neighbours. In particular, we wish to reach a consensus on an alert state when the failure rate exceeds a given threshold. We used a cellular automaton in order to propose solutions in the case of a network with a grid structure. We compared three methods of self-organisation that are partly inspired by physical and biological phenomena. As an application, we envisioned sensor networks or any type of decentralised system with a great number of components.

Concerning the fundamental properties of asynchronous cellular automata, we presented a tutorial on the convergence properties of the 256 Elementary Cellular Automata under the fully asynchronous updating, that is, when only one cell is updated at each time step. We regrouped the results which have been presented in different articles and exposed a full analysis of the behaviour of finite systems with periodic boundary conditions. Our classification relies on the scaling properties of the average convergence time to a fixed point. We presented the different scaling laws that can be found, which fall in one of the following classes: logarithmic, linear, quadratic, exponential and non-converging. The techniques for quantifying this behaviour rely mainly on Markov chain theory and martingales. Most behaviours can be studied analytically but there are still many rules for which obtaining a formal characterisation of their convergence properties is still an open problem.

Our article on the global synchronisation problem was finally published [21]. In this problem, one is asked to find a cellular automaton which has the property that every initial condition evolves into a homogeneous blinking state. We studied this simple inverse problem for the case of one-dimensional systems with periodic boundary conditions. Two paradoxical observations were made: (a) despite the apparent simplicity of finding rules with good statistical results, there exist no perfect deterministic solutions to this problem, (b) if we allow the use of randomness in the local rule, constructing “perfect” stochastic solutions is easy. For the stochastic case, we give some rules for which the mean time of synchronisation varies quadratically with the number of cells and ask if this result can be improved. To explore more deeply the deterministic rules, we code our problem as a SAT problem and use SAT solvers to find rules that synchronise a large set of initial conditions.

7.6. Diagrammatic quantum computing

- Participants: Titouan Carette, Dominic Horsman, Emmanuel Jeandel, Simon Perdrix, Renaud Vilmart.

This year, we have contributed in several ways to the foundations and the applications of the ZX-calculus, a diagrammatic language for quantum computing.

Emmanuel Jeandel, Simon Perdrix, and Renaud Vilmart have introduced a general normal form for ZX-diagrams implying completeness results for various (almost all) fragments of quantum mechanics [28]. Renaud Vilmart has also introduced the simple, meaningful axiomatisation of the full ZX-calculus [31]. This two papers have been published at LICS’19.

Titouan Carette, Emmanuel Jeandel, Simon Perdrix, and Renaud Vilmart, have introduced a new simple categorical construction allowing to deal with non pure quantum evolutions (i.e. involving quantum measurements, discard of quantum systems, and probability mixtures). Wen this new construction coincides with the existing constructions, it provides simpler axiomatisation. For instance, this construction provides a complete equational theory for an extension of the ZX-calculus for arbitrary (non necessary pure) quantum evolutions. This result has been published at ICALP’19 [24].

Titouan Carette, Dominic Horsman (form LIG Grenoble) and Simon Perdrix have provided an axiomatisation for a scalable ZX-calculus where each wire represents a register of qubits, instead of a single qubit in the standard ZX-calculus. The scalable ZX-calculus allows compact representation of quantum algorithms, protocols and quantum codes. This result has been published at MFCS’19 [23]

7.7. Causal Graph Dynamics

- Participants: Pablo Arrighi, Simon Martiel, Simon Perdrix.

Causal Graph Dynamics extend Cellular Automata to arbitrary time-varying graphs of bounded degree. The whole graph evolves in discrete time steps, and this global evolution is required to have a number of symmetries: shift-invariance (it acts everywhere the same) and causality (information has a bounded speed of propagation). Pablo Arrighi (LIS, Marseille), Simon Martiel (Atos-Bull) and Simon Perdrix have considered a natural physics-like symmetry, namely reversibility. In particular, they extended two fundamental results on reversible cellular automata, by proving that the inverse of a causal graph dynamics is a causal graph dynamics, and that these reversible causal graph dynamics can be represented as finite-depth circuits of local reversible gates. These results have been published in the journal Natural Computing [16].

7.8. Contextuality in multipartite pseudo-telepathy graph games

- Participants: Anurag Anshu, Peter Høyer, Mehdi Mhalla, and Simon Perdrix.

Analyzing pseudo-telepathy graph games, Anurag Anshu, Peter Høyer, Mehdi Mhalla, and Simon Perdrix proposed a way to build contextuality scenarios exhibiting the quantum advantage using graph states. A new tool, called multipartiteness width, is introduced to investigate which scenarios are hard to decompose and to show that there exist graphs generating scenarios with a linear multipartiteness width. These results have been published in the Journal of Computer and System Science [15].

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 [20]. 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.

In addition to the natural semantics for MLTS that we proposed in 2018, we also have a small-step operational semantics which gives in particular a fine-grained description of the runtime behavior of the ∇ operator in patterns. It leads in particular to a direct implementation in Lambda-Prolog (which does not contain a native ∇ operator) that allows more expressive constructs (higher-arity types) than our previous presentation.

7.2. Mechanized metatheory revisited

Participant: Dale Miller.

When proof assistants and theorem provers implement the metatheory of logical systems, they must deal with a range of syntactic expressions (e.g., types, formulas, and proofs) that involve variable bindings. Since most mature proof assistants do not have built-in methods to treat bindings, they have been extended with various packages and libraries that allow them to encode such syntax using, for example, De Bruijn numerals. In the paper, [10], Miller puts forward the argument that bindings are such an intimate aspect of the structure of expressions that they should be accounted for directly in the underlying programming language support for proof assistants and not via packages and libraries. He presents an approach to designing programming languages and proof assistants that directly supports bindings in syntax. The roots of this approach can be found in the *mobility* of binders between term-level bindings, formula-level bindings (quantifiers), and proof-level bindings (eigenvariables). In particular, the combination of Church's approach to terms and formulas (found in his Simple Theory of Types) and Gentzen's approach to proofs (found in his sequent calculus) yields a framework for the interaction of bindings with a full range of logical connectives and quantifiers. Miller also illustrates how that framework provides a direct and semantically clean treatment of computation and reasoning with syntax containing bindings.

7.3. New applications of Foundational Proof Certificates

Participants: Kaustuv Chaudhuri, Matteo Manighetti, Dale Miller.

The formal framework of *Foundational Proof Certificates* (FPC) was developed in previous years within the Parsifal team. We continue to push on their applications in a number of settings in computational logic. In 2019, we developed two such new applications.

In order to apply FPCs to the conventional setting of classical logic theorem provers, the FPC setting needed to treat proof evidence containing Skolem functions. Using FPC directly meant that we needed to do such certification without using the mathematical concepts of model-theoretic semantics (i.e., preservation of satisfiability) and choice principles (i.e., epsilon terms). Instead, our proof checking kernel is an implementation of Gentzen's sequent calculus, which directly supports quantifier alternation by using eigenvariables. In [19], we described deskolemization as a mapping from client-side terms, used in proofs generated by theorem provers, into kernel-side terms, used within our proof checking kernel. This mapping which associates skolemized terms to eigenvariables relies on using outer skolemization.

Property-based testing (PBT) is a technique for validating code against an executable specification by automatically generating test-data. In the paper [18], we presented a proof-theoretical reconstruction of this style of testing for relational specifications and employ FPCs to describe test generators. We did this by presenting certain kinds of "proof outlines" that can be used to describe various common generation strategies in the PBT literature, ranging from random to exhaustive, including their combination. We also address the shrinking of counterexamples as a first step towards their explanation. Once generation is accomplished, the testing phase boils down to a standard logic programming search. We could also we lift our techniques to treat data structures containing bindings using λ-*tree* syntax. The λProlog programming language is capable of performing both the generation and checking of tests. We validated this approach by tackling benchmarks in the metatheory of programming languages coming from related tools such as PLT-Redex Property-Based Testing via Proof Reconstruction. This work was done in collaboration with Roberto Blanco, a postdoc from Inria Paris, and Alberto Momigliano, a professor from the University of Milan.

7.4. Historical reflections on proof theory and logic programming

Participant: Dale Miller.

Miller has been working in the area of logic programming and proof theory for more than three decades. Some of his historical reflections on how these two topics influenced each other are contained in the paper [11]. While it is widely known that proof theory has been helpful in shaping the development of logic programming, particular of extensions to conventional Prolog, this paper also documents a few specific examples where logic programming influenced the development of some topics in proof theory.

7.5. Intuitionistic proofs without syntax

Participant: Lutz Straßburger.

We present Intuitionistic Combinatorial Proofs (ICPs), a concrete geometric semantics of intuitionistic logic based on the principles of classical combinatorial proofs. An ICP naturally factorizes into a linear fragment, a graphical abstraction of an IMLL proof net (an arena net), and a parallel contraction-weakening fragment (a skew fibration). ICPs relate to game semantics, and can be seen as a strategy in a Hyland-Ong arena, generalized from a tree-like to a dag-like strategy. Our first main result, Polynomial Full Completeness, is thatICPs as a semantics are complexity-aware: the translations to and from sequent calculus are size-preserving (up to a polynomial). By contrast, lambda-calculus and game semantics incur an exponential blowup. Our second main result, Local Canonicity, is that ICPs abstract fully and faithfully over the non-duplicating permutations of the sequent calculus. These results have been presented at the LICS 2019 conference [23].

7.6. Towards a combinatorial proof theory

Participants: Lutz Straßburger, Benjamin Ralph.

The main part of a classical combinatorial proof is a skew fibration, which precisely captures the behavior of weakening and contraction. Relaxing the presence of these two rules leads to certain substructural logics and substructural proof theory. We investigated what happens if we replace the skew fibration by other kinds of graph homomorphism. This leads us to new logics and proof systems that we call combinatorial. This has been presented at the TABLEAUX 2019 conference [22].

7.7. Combinatorial Proofs for Logics of Relevance and Entailment

Participants: Lutz Straßburger, Matteo Acclavio.

In this work (presented at the WoLLIC 2019 conference [16]) we characterize classical combinatorial proofs which also represent valid proofs for relevant logic with and without the mingle axiom. Moreover, we extend our syntax in order to represent combinatorial proofs for the more restrictive framework of entailment logic.

7.8. On combinatorial proofs for modal logic

Participants: Lutz Straßburger, Matteo Acclavio.

In this work [17], we extend combinatorial proofs to modal logics. The crucial ingredient for modeling the modalities is the use of a self-dual non-commutative operator that has first been observed by Retoré through pomset logic. Consequently, we had to generalize the notion of skew fibration from cographs to Guglielmi's relation webs. Our main result is a sound and complete system of combinatorial proofs for all normal and non-normal modal logics in the S4-tesseract. The proof of soundness and completeness is based on the sequent calculus with some added features from deep inference.

7.9. Deep inference and expansion trees for second-order multiplicative linear logic

Participant: Lutz Straßburger.

In this work, we introduce the notion of expansion tree for linear logic. As in Miller's original work, we have a shallow reading of an expansion tree that corresponds to the conclusion of the proof, and a deep reading which is a formula that can be proved by propositional rules. We focus our attention to MLL2, and we also present a deep inference system for that logic. This allows us to give a syntactic proof to a version of Herbrand's theorem. This has been published in an special issue of MSCS [12].

7.10. A fully labelled proof system for intuitionistic modal logics

Participants: Lutz Straßburger, Marianela Morales.

In this paper we present a labelled sequent system for intuitionistic modal logics such that there is not only one, but two relation symbols appearing in sequents: one for the accessibility relation associated with the Kripke semantics for normal modal logics and one for the preorder relation associated with the Kripke semantics for intuitionistic logic. This puts our system in close correspondence with the standard birelational Kripke semantics for intuitionistic modal logics. As a consequence it can encompass a wider range of intuitionistic modal logics than existing labelled systems. We also show an internal cut elimination proof for our system [30].

7.11. Types by Need

Participants: Beniamino Accattoli, Maico Leberle.

This joint work with Giulio Guerrieri (Post-doc at Bath University) [27] develops a multi type system for call-by-need evaluation of the -calculus. The type system is obtained by combining features by well-known systems for call-by-name and call-by-value. It characterizes termination, and, moreover, its type derivations provide precise information about the number of steps to reach the result. The novelty is that, while the systems for call-by-name and call-by-value are obtained by the linear logic interpretation of these evaluation schemes, call-by-need has no linear logic interpretation.

7.12. Sharing Equality is Linear

Participants: Beniamino Accattoli, Andrea Condoluci, Claudio Sacerdoti Coen.

This work [28] studies how to compare higher-order programs with sharing for sharing equality, that is, for equality of their unshared underlying programs. The point, of course, is to do it efficiently, without unsharing the programs, that would otherwise introduce an exponential blow-up. We develop the first algorithm linear in the size of the shared terms, by adapting the famous Patterson and Wegman algorithm for first-order unification.

7.13. Crumbling Abstract Machines

Participants: Beniamino Accattoli, Andrea Condoluci, Claudio Sacerdoti Coen.

This joint work with Giulio Guerrieri (Post-doc at Bath University) [26] studies a new compilation technique for functional programs, dubbed *crumbling* and resembling the transformation into administrative normal form of Flanagan, Sabry, Duba, and Felleisen. It is shown that it simplifies the design of abstract machines without altering the complexity of the overhead. Moreover, it smoothly scales up to open terms and it does not suffer of the slowdowns of administrative normal forms pointed out by Kennedy.

7.14. Factorization and Normalization, Essentially

Participant: Beniamino Accattoli.

This joint work with Claudia Faggian (CNRS researcher at Paris Diderot) and Giulio Guerrieri (Post-doc at Bath University) [15] refines a rewriting technique for proving factorization and normalization theorems for λ-calculi, that are theorems providing foundations to the design of functional programming languages and proof assistants. We both simplify and extend the scope of a widely used technique by Takahashi. At the concrete level, the new abstract technique is applied to four relevant case studies.

7.15. A Fresh Look at the -Calculus

Participant: Beniamino Accattoli.

This paper [25] is the trace of the invited talk given by Accattoli at FSCD 2019. More than just an abstract, the paper is a lengthy overview of the research on λ-calculus, cost models, sharing, and abstract machines pursued by Accattoli and his co-authors in the last 10 years.

7.16. Abstract Machines for Open Call-by-Value

Participant: Beniamino Accattoli.

This journal paper in collaboration with Giulio Guerrieri (Post-doc at Bath University) [4] outlines a theory of abstract machines for the call-by-value -calculus with open terms. It refines and extends the results by the same authors from 2017, which were among the selected ones from the international conference FSEN 2017 for publication in a journal.

PI.R2 Project-Team

6. New Results

6.1. Effects in proof theory and programming

Participants: Kostia Chardonnet, Emilio Jesús Gallego Arias, Hugo Herbelin, Yann Régis-Gianas, Alexis Saurin, Exequiel Rivas Gadda.

6.1.1. A theory of effects and resources

In collaboration with Thomas Letan (ANSSI), Yann Régis-Gianas developed and proved several properties of a simple web server implemented in Coq using FreeSpec. This work will be presented at CPP 2020.

6.1.2. Call-by-need with probabilistic effects

As a follow up of Chardonnet's Master 1 internship, Kostia Chardonnet and Alexis Saurin continued investigating call-by-need calculi extended with probabilistic choice and started preliminary discussions with Claudia Faggian.

6.1.3. Proof-search, algebraically and graphically

Alexis Saurin worked on proof search in a proof-net scenario, that is proof-net search. A key aspect of proof construction is a management of non-determinism in bottom-up sequent-proof construction, be it when the search succeeds or when facing a failure and the need for backtracking. This is partially dealt with by focussing proof-construction, which reduces drastically the search space while retaining completeness of the resulting proof space (both at the provability level and at the denotational level).

His approach consists in viewing proof-search and sequentialisation as dual aspects of partial proof structures (that is proof nets with open premisses). In particular, he builds on Lafont's parsing criterion to obtain a proof-construction algorithm in which the proof space is not a search tree, as in sequent-calculus, but a dag allowing to share proof-construction paths.

Emilio Jesús Gallego Arias collaborates with Jim Lipton from Wesleyan University on the development of algebraic models for proof search.

6.2. Reasoning and programming with infinite data

Participants: Kostia Chardonnet, Lucien David, Abhishek De, Farzad Jafar-Rahmani, Luc Pellissier, Yann Régis-Gianas, Alexis Saurin.

This theme is part of the ANR project Rapido (see the National Initiatives section) which ended octobre 1st 2019.

6.2.1. Proof theory of non-wellfounded and circular proofs

6.2.1.1. Validity conditions of infinitary and circular proofs

In collaboration with David Baelde, Amina Doumane 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, Alexis Saurin studied together with David Baelde, Amina Doumane and Denis Kuperberg a relaxed validity condition to allow infinite branches to be supported by threads which may leave the infinite branch, visiting other parts of the proofs and 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 in the sequent calculus due to a discrepancy between the sequential nature of proofs in sequent calculus and the parallel nature of threads. David Baelde, Amina Doumane, Denis Kuperberg and Alexis 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.

Several directions of research have therefore been investigated from that point:

- Decidability can be recovered by constraining the shapes of bounces (bounding the depth of bounces). They actually exhibited a hierarchy of criteria, all decidable and satisfying the fact that their union corresponds to bouncing validity (which is therefore semi-decidable)
- While the result originaly held only for the multiplicative fragment of linear logic, the result was extended to multiplicative and additive linear logic.

Those results are currently submitted.

6.2.1.2. On the complexity of the validity condition of circular proofs

Alexis Saurin, together with Rémi Nollet and Christine Tasson, characterised the complexity of deciding the validity of circular proofs. While deciding validity was known to be in PSPACE, they proved that, for $\mu MALL$ proof, it is in fact a PSPACE-complete problem.

The proof is based on a deeper exploration of the connection between thread-validity and the size-change termination principle, a standard tool to prove program termination.

This result has been presented and published at TABLEAUX 2019 [41].

6.2.1.3. Proof nets for non-wellfounded proofs

Abhishek De and Alexis Saurin set the basis of the theory of non-wellfounded and circular proofs nets (in the multiplicative setting). Non-wellfounded proof nets, aka infinets, were defined extending Curien's presentation of proof nets allowing for a smooth extension to fixed point logics. The aim of this work is to provide a notion of canonical proof obiects for circular proofs free from the irrelevant details of the syntax of the sequent calculus. The first results were published in TABLEAUX 2019 [38] and provide a correctness condition for an infinet to be sequentialisable in a sequent proof.

The results of the TABLEAUX paper are limited in that they only address the case of proofs with finitely many cuts inferences. Abhishek De and Alexis Saurin are currently investigating, with Luc Pellissier, the general case of infinitely many cut in order to then lift the results from straight thread validity to bouncing thread validity.

6.2.2. On the denotational semantics of non-wellfounded proofs

Farzad Jafar-Rahmani started his PhD under the supervision of Thomas Ehrhard and Alexis Saurin in October 2019. His PhD work will focus on the denotational semantics of circular proofs of linear logic with fixed points. After working on the denotational semantics of finitary proofs for linear logic with fixed points (with Kozen rules) during his master, he is currently working at understanding the denotational counterpart of the validity condition of circular proofs.

6.2.3. Towards inductive and coinductive types in quantum programming languages

Kostia Chardonnet started his PhD under the supervision of Alexis Saurin and Benoît Valiron in November 2019. Previously, he did his MPRI Master internship under their joint supervision on designing a calculus of reversible programs with inductive and coinductive types. His research focused on extending a languages of type isomorphisms with inductive and coinductive types and understanding the connections of those reversible programs with $\mu MALL$ type isomorphisms and more specifically with $\mu MALL$ focused circular proof isomorphisms. In his PhD, he shall extend this to the case of a quantum programming language with inductive and coinductive data types.

6.2.4. Theory of fixed points in the lambda-calculus

The results of Alexis Saurin in collaboration with Giulio Manzonetto, Andrew Polonsky and Jacob Grue Simonsen, on two long-standing conjectures on fixed points in the λ-calculus – the "fixpoint property" and the "double-fixpoint conjecture" – have now appeared in the Journal of Logic and Computation [34]. The former asserts that every λ-term admits either a unique or an infinite number of β-distinct fixpoints while the second, formulated by Statman, says that there is no fixpoint satisfying $Y\delta = Y$ for $\delta = \lambda y, x.x(yx)$. They proved the first conjecture in the case of open terms and refute it in the case of sensible theories (instead of β). Moreover, they provide sufficient conditions for both conjectures in the general case. Concerning the double-fixpoint conjecture, they propose a proof technique identifying two key properties from which the results would follow, while they leave as conjecture to prove that those actually hold.

6.3. Effective higher-dimensional algebra

Participants: Antoine Allioux, Pierre-Louis Curien, Alen Durić, Eric Finster, Yves Guiraud, Amar Hadzihasanović, Cédric Ho Thanh, Matthieu Sozeau.

6.3.1. Rewriting methods in higher algebra

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 [64] 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 [63] and Poincaré-Birkhoff-Witt bases [100]. Then, they transposed the construction of [14], based on an extension of Squier's theorem [104] 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 was published in Mathematische Zeitschrift [32].

Yves Guiraud has written and defended his "Habilitation à diriger des recherches" manuscript, as a survey on rewriting methods in algebra based on Squier theory [26]. The defense was held in June 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).

Building on [9], Yves Guiraud is currently finishing with Matthieu Picantin (Univ. Paris Diderot) a work that generalises already known constructions such as the bar resolution, several resolutions defined by Dehornoy and Lafont [73], and the main results of Gaussent, Guiraud and Malbos on coherent presentations of Artin monoids [11], 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.

Yves Guiraud has started a collaboration with Najib Idrissi (IMJ-PRG, Univ. Paris Diderot) whose aim is to understand the relation between several different methods known to compute small resolutions of algebras and operads: those based on rewriting methods (Anick, Squier) and those that stem from Koszul duality theory.

6.3.2. Normalisation of monoids

Alen Durić started his Phd thesis (supervised by Yves Guiraud and Pierre-Louis Curien) in October 2019. His work so far has been mostly bibliographical. The goal is to combine methods from rewriting theory (and in particular the method of homotopical completion and reduction developed by Guiraud-Malbos-Mimram) and methods developed by Dehornoy and his coauthors in the study of monoids with Garside families, and by Dehornoy-Guiraud in the study of normalisation for monoids. Alen Durić is currently experimenting with some examples taken from these latter works, with the goal of building coherent presentations for them using the former methods.

6.3.3. Topological aspects of polygraphs

Amar Hadzihasanović joined the team at the end of November 2019, as a one-year postdoc funded by FSMP. He has been working intensively on the study of shapes appropriate for the description of higher cells as needed in various approaches to higher categories and higher structures. Amar Hadzihasanović's project is to recast his ideas in the framework of polygraphs, with the aim of bringing topological insights into the study of higher-dimensional rewriting.

6.3.4. Opetopes

The work of Pierre-Louis Curien, Cédric Ho Thanh and Samuel Mimram on syntactic and type-theoretic presentations of opetopes and opetopic sets has been submitted to a journal, and a short version has been presented at the LICS conference in Vancouver this year [45].

Cédric Ho Thanh, in collaboration with Chaitanya Leena Subramaniam, has defined the notion of "opetopic algebras" that leverages the subtle combinatorics of opetopes. This framework encompasses categories, planar operads, and Loday's combinads over planar trees. They have defined an opetopic nerve functor that fully embeds each category of opetopic algebras into the category of opetopic sets. In particular, they obtain fully faithful opetopic nerve functors for categories and for planar coloured operads. These results have been written up in [51]. This work is the first in a series aimed at using opetopic spaces as models for higher algebraic structures. In particular, the aim is to provide new models for infinity-categories and infinity-operads.

6.3.5. Foundations and formalisation of higher algebra

Antoine Allioux (PhD started in February 2018), 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. During the first year of Allioux's PhD, they studied 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 [90]. They later realised that this internalisation is however incoherent as presented in pure type theory, essentially because of its reliance on equality types. Since then, they switched to a different view, describing opetopes as an external construction and relying on strict equalities in the metatheory to avoid the coherence problem. Opetopic type theory should then be, similarly to cubical type theory, a type theory indexed over these opetopic structures, where grafting and substitution are computional operations. They are now concentrating on showing that the modified inductive characterisation of opetopes and their algebras, still definable in type theory, gives rise to the standard notion of opetopes in mathematics, an original result in itself.

6.4. Incrementality

Participants: Thibaut Girka, Yann Régis-Gianas.

In collaboration with Paolo Giarrusso (EPFL, Switzerland), Philipp Shuster (Univ. of Tübingen, Germany), 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 presented at ESOP 2019 in Prague. Yann Régis-Gianas also presented this work at several places (Gallium seminar, Galinette seminar, and Chocola seminar).

In collaboration with Olivier Martinot (Paris Diderot), Yann Régis-Gianas studied a new technique to implement incrementalised operations on lists.

In collaboration with Faridath Akinotcho (Paris Diderot), Yann Régis-Gianas studied an incrementalisation of the Earley parsing algorithm.

6.5. Metatheory and development of Coq

Participants: Félix Castro, Emilio Jesús Gallego Arias, Gaëtan Gilbert, Hugo Herbelin, Pierre Letouzey, Cyprien Mangin, Thierry Martinez, Yann Régis-Gianas, Matthieu Sozeau, Théo Winterhalter, Théo Zimmermann.

6.5.1. Meta-programming and Metatheory of Coq

The MetaCoq project started last year, providing the means to program program transformations and general purpose plugins in Coq, using a quoting/unquoting mechanism. This year, they extended the framework to specify the theory, including the reduction, cumulativity and typing relations of the Polymorphic, Cumulative Calculus of Inductive Constructions at the basis of Coq. Matthieu Sozeau, together with Simon Boulier, Nicolas Tabareau and Théo Winterhalter at Galinette, Cyril Cohen at Marelle, Yannick Forster and Fabian Kunze at the University of Saarbrucken and Abhishek Anand and Gregory Malecha at BedRock Systems, Inc co-authored [54] a full description of the resulting theory (to appear in JAR). This allows for the verification of term manipulations with respect to typing: syntactic translations but also reflexive tactics glue code can hence be verified. The article also develops an alternative extraction mode to OCaml allowing the efficient compilation and execution of meta-programs written in the Template Monad. An example partial extraction of Coq programs to call-by-value pure lambda-calculus is developed this way.

Following up on this work, Matthieu Sozeau led a metatheoretical study of Coq in Coq, proving the basic metatheoretical properties of the typing relation, and developed together with Yannick Forster (Saarbrucken) and Simon Boulier, Nicolas Tabareau and Théo Winterhalter (Gallinette) verified correct versions of type-checking and erasure for a large subset of Coq. This work involved the production of a fully-precise specification for the type theory implemented by Coq, cleaning up the previously untested typing specification, and variants of the algorithms used in its kernel ammenable to proofs of correctness. The corresponding implementations can be extracted and provide an alternative, verified checker for Coq terms, that can run on medium-sized examples. This work will be presented [35] at POPL in New Orleans in January 2020.

6.5.2. Homotopy type theory

Hugo Moeneclaey started in September 2019 a PhD on the syntax of spheres in homotopy type theory, under the supervision of Hugo Herbelin.

Hugo Herbelin and Hugo Moeneclaey worked on the syntax of a variant of Cohen, Coquand, Huber and Mörtberg's Cubical Type Theory justified by an iterated parametricity model where equality on types is defined to be equivalence of types, thus satisfying univalence by construction.

6.5.3. Computational contents of the axiom of choice

Hugo Herbelin developed in collaboration with Nuria Brede (U. Potsdam) a unified logical structure for choice and bar induction principles.

6.5.4. Computational contents of Gödel’s constructible universe

Félix Castro started his PhD under the supervision of Hugo Herbelin and Alexandre Miquel in September 2019. His PhD work will focus on the computational contents of Gödel’s constructible universe. Previously, he worked on the formalisation of the ramified analytical hierarchy in classical second-order arithmetic.

6.5.5. Dependent pattern-matching and recursion

Together with Cyprien Mangin, Matthieu Sozeau refined the treatment of dependent pattern-matching in the Equations plugin. By carefully studying the type of equalities between indexed inductive types, he devised a new criterion for the elimination of equalities between inductive families based on the notion of forced arguments of constructors, resulting in a simplification of the setup of Cockx and Devriese [68] for simplification of dependent pattern-matching without K. This improved simplifier is part of the latest version of the Equations plugin, which also provides better support for the definition of mutual and well-founded recursive definitions on indexed inductive types. This work was presented at ICFP 2019 in Berlin [36]. A longer journal version is in preparation, along with a dedicated tutorial on Equations slated for inclusion in a new volume of the Software Foundations series dedicated to advanced tools.

Thierry Martinez continued part time 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.

6.5.6. 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 analysed 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 quantitative results were completed with qualitative data coming from interviews with main Coq developers, which help interpret them. They validate *a posteriori* the usefulness of such a switch. A paper [43] has been published at ICSME 2019, which is the leading conference on the topic of Software Maintenance and Evolution.

Besides, Théo Zimmermann also studied and influenced the pull-based model that is now used for the development of Coq, he improved the release management process and tools, he studied package distribution and maintenance, in particular with the foundation of the coq-community organisation in 2018, which has taken off by attracting 19 maintainers, and hosting 25 projects. All of these topics are presented in the PhD thesis [28] that he defended in December 2019.

Emilio J. Gallego Arias and Théo Zimmerman took the roles of release managers for the Coq 8.12 and will oversee this release, planned for mid-2020.

Emilio J. Gallego Arias and Théo Zimmerman discussed on future plans for compositional proof checking using the Dune build system, which will include a new library format for Coq. The Dune team was informed, with Emilio J. Gallego Arias participating in the bi-weekly developer meetings. Emilio J. Gallego Arias also started discussion with the Debian OCaml maintainers (who are located at IRIF) as to see how to better integrate Dune with the Debian packaging workflow.

Emilio J. Gallego Arias designed the Coq instrumentation used in the [103] paper, which collects and analyses changes to proof scripts.

Emilio J. Gallego Arias and Karl Palmskog released a new version of the Coq SerAPI tool, which has been used in some recent proof engineering efforts, such as [65], the machine-learning environments CoqGYM and Proverbot9001 [108], [55], offering state of the art proof automation after training with proof data sets, and the educational user interface WaterProof [56]. SerAPI has also been used in some other works undergoing review and thus yet not public.

Emilio J. Gallego Arias and Shachar Itzhaky released a new version of the educational Coq frontend jsCoq [10], and assisted a few users who have been preparing courses using it.

Emilio J. Gallego Arias maintains an ongoing collaboration with the Deducteam group at Inria Saclay on the topic of interactive proof methods and standards; this has resulted in the release of an experimental LSP server for the Lambdapi theorem prover.

Emilio J. Gallego Arias, Hugo Herbelin, and Théo Zimmerman participate in the Logipedia project led by Gilles Dowek, which aims to develop a standard proof interchange format.

6.5.7. Software Infrastructure

Emilio J. Gallego Arias did significant work to refactor the Coq codebase in preparation for further work on incremental and multi-core aware type checking.

6.5.8. Dissemination activities

Emilio J. Gallego Arias and Théo Zimmerman organised the Coq meetup, an after-work event targeting industry and other communities outside academia.

6.5.9. Coordination of the development of Coq

Hugo Herbelin, Matthieu Sozeau, Emilio J. Gallego Arias 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.

6.6. Formalisation and verification

Participants: Pierre-Louis Curien, Lucien David, Emilio Jesús Gallego Arias, Kailiang Ji, Pierre Letouzey, Jean-Jacques Lévy, Cyprien Mangin, Daniel de Rauglaudre, Yann Régis-Gianas, Alexis Saurin, Matthieu Sozeau.

6.6.1. Proofs and surfaces

The joint work of Pierre-Louis Curien with Jovana Obradović (former PhD student of the team and now postdoc in Prague), 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, has been submitted to a journal, and has been presented at the conference Topology, Algebra, and Categories in Logic (TACL) 2019, Nice, in June 2019 [53].

6.6.2. A Coq formalisation of the first-order predicate calculus

In relation with a logic course for master students, Pierre Letouzey made a Coq formalisation of the first-order predicate calculus. The logical rules are expressed in a natural deduction style (with explicit contexts). Pierre Letouzey proposed two low-level representations of formulas : one based on quantifiers with names, the other using "locally nameless" techniques. The equivalence between the two settings has been proved correct. Using this deep embedding, Pierre Letouzey formalised in Coq the whole course notes (prepared some years ago by Alexandre Miquel), including the completeness theorem for this logic. This development is available at https://gitlab.math.univ-paris-diderot.fr/letouzey/natded.

6.6.3. A Coq formalisation of circular proofs and their validity condition

During the summer 2019, Alexis Saurin supervised Lucien David's M1 internship on formalizing in Coq circular proofs and their meta-theory. This work built on Xavier Onfroy's previous work as well as on Pierre Letouzey's formalisation of the predicate calculus in natural deduction mentioned above. While the previous work by Xavier Onfroy was both contributing to the proof theory part and the ω-automata part (which is need for the decidability theorem), Lucien David completely focused on the the proof theory side. In particular, he was able to improve significantly on Xavier Onfroy's formalisation by using ideas from Letouzey's formalisation of natural deduction and by interacting with Pierre Letouzey and Alexis Saurin. This development is available at https://github.com/LuluDavid/CircularProofsValidity.

6.6.4. Lexing and regular expressions in Coq

Pierre Letouzey and Yann Régis-Gianas revisited in Coq classical techniques about lexing and regular expressions. In particular, regular expressions (with complement and conjunction) have been formalised, as well as their Brzozowski derivatives, and the finiteness theorem due to Brzozowski : a given regular expression admits only a finite number of derivatives (up to some equivalence). Both the general equivalence (based on language identity) and practical approximations (similarities) has been considered (and proved decidable). From that, the algorithms building recognizing automata (with derivatives as states) have been formalised and proved, leading to the minimal automata when using the general equivalence (but at a high cost), or to practical approximations of the minimal automata when using various similarities. This work is still ongoing. For instance, the correctness proof of a particular similarity used in an existing implementation (ml-ulex) is quite elusive for the moment. They also plan to extend this development up to a full-scale tool a la ocamllex in Coq.

6.6.5. Real Numbers as sequences of digits in Coq

Daniel de Rauglaudre has been continuing the formalisation of real numbers defined as sequences of digits in any radix with the LPO axiom/oracle (Limited Principle of Omniscience). Although the operations (additions and multiplications) work with this method, the proof of associativity of addition needs more work to be achieved. This development is available at https://github.com/roglo/coq_real/.

6.6.6. Category theory in Coq

Daniel de Rauglaudre started an implementation in Coq of Category theory in Coq, using in particular theorems coming from HOTT (HOmotopy Type theory) that he implemeted some years ago. Several notions around Categories have been defined. For example, Yoneda Lemma, among others. This development is available at https://github.com/roglo/mycoqhott/.

6.6.7. Number theory in Coq

Daniel de Rauglaudre started and almost completed the formalisation in Coq of the proof of Euler's Product Formula, stating that the Riemann zeta function, which is a sum on all the natural numbers, is also a product on all the prime numbers. He also added several theorems about the prime numbers. This development is available at https://github.com/roglo/coq_euler_prod_form.

6.6.8. Proofs of algorithms on graphs

Jean-Jacques Lévy and Chen Ran (a PhD student at the Institute of Software, Beijing) pursued their work about formal proofs of graph algorithms. Their goal is to provide proofs of algorithms checked by computer and human readable. In 2019, they presented at ITP 2019 a joint paper with Cyril Cohen, Stephan Merz and Laurent Théry on this work [37]. This article compared formal proofs in three different systems (Why3, Coq, Isabelle/HOL) of Tarjan (1972) linear-time algorithm computing the strongly connected components in directed graphs.

The current work is to have a proof of the implementation of this algorithm with imperative programming and memory pointers. They also planed to produce formal proofs of other abstract algorithms such as the Hopcroft-Tarjan (1972) linear-time algorithm for planarity testing in undirected graphs.

6.6.9. Certified compilation and meta-programming

Matthieu Sozeau participates to the CertiCoq project led by Andrew Appel at Princeton (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. Together with Yannick Forster at the University of Saarbrucken and the MetaCoq team, Matthieu Sozeau focused the verification of type-checking and erasure which were previously trusted parts of the system. The new verified erasure function fills a gap in the proof of correctness of compilation from Gallina terms down to C-light. The whole compiler can be run on realistic examples (the erasure phase does take most of the compilation time and should be optimised further).

In collaboration with Xavier Denis (Paris Diderot), Yann Régis-Gianas formalised and built a compiler for Mtac2. A paper is in preparation.

STAMP Project-Team

6. New Results

6.1. Hol-Light and Elpi

Participants: Enrico Tassi, Marco Maggesi [University of Florence, Italy].

We implemented an elaborator for HOL-Light in Elpi. In particular the new elaborator supports coercions and overloaded notations for algebraic structures.

6.2. Generating equality tests for inductive types

Participant: Enrico Tassi.

We show how to derive in a modular fashion equality tests for a wide variety of inductive type definitions. This makes an instrumental use of parametricity. This work has been published in an international conference [10]. This is also an interesting case study for the Elpi language [2].

6.3. Re-designing the state machine of Coq

Participants: Enrico Tassi, Maxime Dénès.

We redesigned the state machine of Coq to improve its support for LSP-based user interfaces [0]. In particular, we decoupled the representation of the document as seen by the User-Interface and the structured document as used by the STM to decide what to compute and how (in which order).

6.4. Formal proofs on session types

Participants: Enrico Tassi, Cinzia Di Giusto [University of Nice], Marco Giunti [New University of Lisbon], Kirstin Peters [University of Darmstadt], Antonio Ravara [New University of Lisbon].

We formalized in Coq and Isabelle a linear, monadic Pi calculus, its labelled transition system, and type system. We proved the properties of subject reduction and absence of linearity violation. These are based on De Bruijn levels and Nominals with the objective of comparing the approaches and provide automation for recurrent goals. This is done in Project PROSE (Provers for Sessions).

6.5. Formal proofs of an axiomatization of graphs with tree-width two

Participants: Christian Doczkal, Damien Pous [CNRS, ENS de lyon].

We finished the formalization of a completeness proof for an axiomatization of graphs of treewidth at most two in Coq+MathComp. This work was submitted for publication in a conference [11]. We are also revising the article presenting our proof of the Minor-Exclusion property for Treewidth-Two graphs [15] for publication in a journal. Most of the formal proofs are available from the following web-site https://perso.ens-lyon.fr/damien.pous/covece/graphs/.

6.6. Formal study of double-word arithmetic algorithms

Participants: Laurence Rideau, Jean-Michel Muller [CNRS, ENS de Lyon].

We finished the formalization of double-word arithmetic algorithms, described in the article *Tight and rigourous error bounds for basic building blocks of double-word arithmetic* [16].

[0] LSP stands for Language Server Protocol

Thanks to the formalization, errors were found in the proofs, but the stated results (correction of algorithms and error limits) were proven correct. On the other hand, for the purposes of this formalization, we had to develop a more general version of the proof of the Fast2Sum algorithm, which should soon be integrated into the Flocq library.

An article describing this work of formalisation is being written.

6.7. Approximations using Chebyshev polynomials

Participants: Laurent Théry, Florian Steinberg [Inria Saclay, Toccata project-team].

Florian Steinberg and Laurent Théry have been working on polynomial approximations using Chebyshev polynomials. This works has been presented at the ANR FastRelax final meeting (Lyon, June 2019) and is available as a library at https://github.com/FlorianSteinberg/Cheby.

6.8. Formalizing computational analysis

Participants: Laurent Théry, Florian Steinberg [Inria Saclay, Toccata project-team], Holger Thies [Kyushu University, Fukuoka].

Florian Steinberg, Holger Thies, and Laurent Théry have been working on formalizing computational analysis. This work is described in a paper to be submitted for publication [13]. A shorter version was published in a conference [9].

6.9. Formal study of probabilistic programs

Participants: Cécile Baritel-Ruet, Benjamin Grégoire, José Bacelar Almeida [INESC TEC], Manuel Barbosa [INESC TEC], Gilles Barthe [IMDEA], Sonia Belaïd [CryptoExpert], Matthew Campagna [AWS], Gaëtan Cassiers [UCL], Sunjay Cauligi [UC San Diego], Ernie Cohen [AWS], François Dupressoir [University of Surrey], Pierre-Alain Fouque [Université Rennes 1], Charlie Jacomme [LSV], Steve Kremer [Inria Nancy Grand-Est, PESTO project Team], Adrien Koutsos [LSV], Vincent Laporte [Inria], Tiago Oliveira [INESC TEC], Vitor Pereira [INESC TEC], Bernardo Portela [INESC TEC], Alley Stoughton [Boston University], François-Xavier Standaert [UCL], Deian Stefan [UC San Diego], Pierre-Yves Strub [Ecole Polytechnique], Serdar Tasiran [AWS].

We provide two differents tools:

- EasyCrypt (see http://www.easycrypt.info/) is a toolset for reasoning about relational properties of probabilistic computations with adversarial code. Its main application is the construction and verification of game-based cryptographic proofs.
- Jasmin (see https://github.com/jasmin-lang/jasmin) is certified compiler to generate high-speed and high-assurance cryptographic code.

6.10. Security of a key management service

Participants: Benjamin Grégoire, José Bacelar Almeida [INESC TEC], Manuel Barbosa [INESC TEC], Gilles Barthe [IMDEA], Matthew Campagna [AWS], Vitor Pereira [INESC TEC], Bernardo Portela [INESC TEC], Pierre-Yves Strub [Ecole Polytechnique], Serdar Tasiran [AWS].

We have developed 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 AWS KMS; it governs the toplevel keys that anchor the security of encryption services at AWS. We show that the protocol securely implements an ideal distributed encryption mechanism under standard cryptographic assumptions. The proof is machine-checked in the EasyCrypt proof assistant and is the largest EasyCrypt development to date. This work corresponds to a contract with AWS and has been published in a major computer security conference [3].

6.11. High-assurance and high-speed SHA-3

Participants: Cécile Baritel-Ruet, Benjamin Grégoire, José Bacelar Almeida [INESC TEC], Manuel Barbosa [INESC TEC], Gilles Barthe [IMDEA], François Dupressoir [University of Surrey], Vincent Laporte [Inria], Tiago Oliveira [INESC TEC], Alley Stoughton [Boston University], Pierre-Yves Strub [Ecole Polytechnique].

We have developed a high-assurance and high-speed implementation of the SHA-3 hash function. Our implementation is written in the Jasmin programming language, and is formally verified for functional correctness, provable security and timing attack resistance in the EasyCrypt proof assistant. Our implementation is the first to achieve simultaneously the four desirable properties (efficiency, correctness, provable security, and side-channel protection) for a non-trivial cryptographic primitive. Concretely, our mechanized proofs show that:

1. The SHA-3 hash function is indifferentiable from a random oracle, and thus is resistant against collision, first and second preimage attacks;
2. The SHA-3 hash function is correctly implemented by a vectorized x86 implementation.

Furthermore, the implementation is provably protected against timing attacks in an idealized model of timing leaks. The proofs include new EasyCrypt libraries of independent interest for programmable random oracles and modular indifferentiability proofs. This work has been published at an international conference [4].

6.12. A domain-specific language for timing sensitive computation

Participants: Benjamin Grégoire, Sunjay Cauligi [UC San Diego], Gilles Barthe [IMDEA], Deian Stefan [UC San Diego].

Real-world cryptographic code is often written in a subset of C intended to execute in constant-time, thereby avoiding timing side channel vulnerabilities. This C subset eschews structured programming as we know it: if-statements, looping constructs, and procedural abstractions can leak timing information when handling sensitive data. The resulting obfuscation has led to subtle bugs, even in widely-used high profile libraries like OpenSSL. To address the challenge of writing constant-time cryptographic code, we have participate to the development of FaCT, a crypto DSL that provides high-level but safe language constructs. The FaCT compiler uses a secrecy type system to automatically transform potentially timing-sensitive high-level code into low-level, constant-time LLVM bitcode. While the language and the type system has been developed by our collaborator, we have formalized the constant-time transformation. We have performed an empirical evaluation that uses FaCT to implement core crypto routines from several open-source projects including OpenSSL, libsodium, and curve25519-donna. Our evaluation shows that FaCT's design makes it possible to write readable, high-level cryptographic code, with efficient, constant-time behavior. This work has been published at an international conference [7].

6.13. Proving equivalence between probabilistic programs

Participants: Benjamin Grégoire, Gilles Barthe [IMDEA], Steve Kremer [Inria Nancy Grand-Est, PESTO project Team], Pierre-Yves Strub [Ecole Polytechnique].

We have developed principled methods for proving equivalence between probabilistic programs that operate over finite fields and related algebraic structures. We have focused on three essential properties: program equivalence, information flow, and uniformity. We give characterizations of these properties based on deducibility and other notions from symbolic cryptography. We use (sometimes improve) tools from symbolic cryptography to obtain decision procedures or sound proof methods for program equivalence, information flow, and uniformity. A partial implementation of our approach is integrated in EasyCrypt and in MaskVerif. This work has been published at an international conference [6].

6.14. MaskVerif: automated verification of higher-order masking in presence of physical defaults

Participants: Benjamin Grégoire, Gilles Barthe [IMDEA], Sonia Belaïd [CryptoExpert], Gaëtan Cassiers [UCL], Pierre-Alain Fouque [Université Rennes 1], François-Xavier Standaert [UCL].

Power and electromagnetic based side-channel attacks are serious threats against the security of cryptographic embedded devices. In order to mitigate these attacks, implementations use countermeasures, among which masking is currently the most investigated and deployed choice. Unfortunately, commonly studied forms of masking rely on underlying assumptions that are difficult to satisfy in practice. This is due to physical defaults, such as glitches or transitions, which can recombine the masked data in a way that concretely reduces an implementation's security. We have developed and implemented an automated approach for verifying security of masked implementations in presence of physical defaults (glitches or transitions). Our approach helps to recover the main strengths of masking: rigorous foundations, composability guarantees, automated verification under more realistic assumptions. This work contributes to demonstrate the benefits of language-based approaches (specifically probabilistic information flow) for masking. This work was published at an international conference [5].

6.15. Frame type theory

Participants: Cyril Cohen, Assia Mahboubi [Inria Rennes Bretagne Atlantique, Gallinette project-team], Xavier Montillet [University of Nantes].

Writing modular programs in proof assistants is notoriously difficult. A significant literature and implementation effort is devoted to the topic, with approaches ranging from adding new constructions to the underlying logic, to adding features to the proof assistant. However, all current options (including records, sections and modules) are unsatisfactory in one way or another. In this work in progress we aim at reconciling several options using frames. The central idea is to consider records where some fields do not have a value yet. We will call these generalized records frames, and will say that a field is a definition (resp. abstraction) if it has (resp. does not have) a value. Frames can also be thought of as a reification of the contexts of CiC, as presented in the Coq manual.

6.16. Automated refinements on algorithms in Lean

Participants: Cyril Cohen, Tobias Grosser [ETH Zurich], Utz Haus [CRAY EMEA Research Lab], Chris Hughes [Imperial college].

We have experimented with Applying manual and automated program refinements techniques to a simple algorithm, in Lean, with Tobias Grosser, Utz Haus and Chris Hughes, in Zürich. Experiments on this topic are available at the following address https://github.com/ChrisHughes24/LP.

This work also includes investigations on parametricity in Lean as visible at the following address https://github.com/CohenCyril/mathlib/tree/param.

6.17. Parametricity in Template Coq

Participants: Cyril Cohen, Damien Rouhling, Assia Mahboubi [Inria Rennes Bretagne Atlantique, Gallinette project team], Nicolas Tabareau [Inria Rennes Bretagne Atlantique, Gallinette project team].

We study the implementation of parametricity in Template Coq and improve on the work proposed the article *Equivalence for free!* [17]. This work is available at https://github.com/CoqHott/parametricity-a-la-carte.

6.18. A hierarchy builder

Participants: Kazuhiko Sakaguchi, Cyril Cohen.

We are studying how to generate mathematical structures from their axioms using the high-level language provided by the Coq-Elpi experiment. Ongoing experiments are visible at the following address https://github.com/math-comp/hierarchy-builder.

6.19. Adding measure theory to mathematical components analysis

Participants: Cyril Cohen, Damien Rouhling, Laurence Rideau, Reynald Affeldt [AIST, Japan], Georges Gonthier [Inria Saclay Ile de France, Specfun project team], Marie Kerjean [Inria Rennes Bretagne Atlantique, Gallinette project team], Assia Mahboubi [Inria Rennes Bretagne Atlantique, Gallinette project team], Pierre-Yves Strub [Ecole Polytechnique].

We started extending mathematical components analysis [14] with measure theory and Lebesgue-Stieljes integral. We are taking inspiration from work done on Coquelicot and in the MILC project (DIM-RFSI).

6.20. A formal description of exact real arithmetic

Participants: Yves Bertot, Nicolas Magaud [University of Strasbourg].

We revisited an old package available in the contributions to the Coq system, where algorithms to perform real number computations were described. This package was using primitives described using axioms. We showed that these axioms were faulty and proposed solutions to salvage the package and make it more safely usable in the future.

6.21. Formal study of a triangulation algorithm

Participant: Yves Bertot.

We wish to describe a triangulation algorithm in a way that respects both a high level of abstraction and a precise account of pointer manipulations. Using refinements approaches as in CoqEAL, we hope that this can lead to efficient implementation that are derived from the formal description.

6.22. Formal study of Voronoi diagrams and Fortune's algorithm

Participants: Ahmed Khulaif A Alharbi, Yves Bertot.

Voronoi diagrams are an example of data that can be used to solve problems in robot motion planning. In this experiment, we provided a formal description of Fortune's algorithm to compute such diagrams, together with a framework to animate this algorithm. Formal proofs of correctness will be the next step.

6.23. Formal study of a cell-decomposition algorithm

Participants: Julien Lamiroy, Yves Bertot.

To solve robot motion planning problems, a simple approach is to decompose the available space into obstacle-free cells and to more from one cell to another only by boundaries that are also obstacle free. We developed a formal description of an algorithm producing this kind of decomposition, with the aim of providing formal proofs of correctness in the long run.

6.24. A guide to use Coq for security evaluations

Participants: Maxime Dénès, Yves Bertot, Vincent Laporte, Arnaud Fontaine [ANSSI], Thomas Letan [ANSSI].

Common Criteria are an international standard for computer security certification. Evaluations are rated with Evaluation Assurance Levels, from 1 to 7. Eal6 and EAL7 require developers to conduct a formal analysis of their product with respect to certain security properties.

In France, the Certification Body (the entity emitting Common Criteria certificates) is part of the ANSSI (*l'Agence Nationale de la Sécurité des Systèmes d'Information*, also referred to as the French Cybersecurity Agency), and is one of the few emitters of EAL6 and EAL7 certificates.

Coq has already been used to support Common Criteria formal analysis. The ANSSI and Inria have been collaborating on an authoritative document to introduce guidelines and rules for formal analyses supported by Coq, in order to make these developments easier to read and evaluate.

6.25. Formalization of the Poincaré disk model in Isabelle

Participants: Pierre Boutry, Danijela Simić [University of Belgrade], Filip Marić [University of Belgrade].

The Poincaré disk model is a model that can be shown to satisfy all axioms of Tarski's system of geometry at the exception of the parallel postulate. We developed a formal proof of this fact in the Isabelle system and submitted an article for publication. Reviewers suggested that we add a proof that the postulate of the existence of limiting parallels does hold. This completes neatly the work on this topic, as it allows us to exhibit that the Poincaré disk model is not only a counter-model for the parallel postulate but also a model of hyperbolic geometry. An improved version of the article will be submitted soon.

6.26. Integration of the GeoCoq library to Logipedia

Participants: Pierre Boutry, Gaspard Ferey [Inria Saclay Ile de France, Deducteam project team].

We have proofs of independence of the parallel postulate for several models of hyperbolic geometry (among which the Poincaré disk model). An objective is to provide formal proofs that these models are actually isomorphic. An issue for this objective is the question of re-usability, because the formal proofs that we have so far exist in the realms of different theorem provers. The Logipedia effort is an attempt to make proofs from different proofs systems work together, by using a tool called Dedukti as a go-between. A particular point is to be able to translate proofs already done in Coq, namely the GeoCoq library, into proofs verifiable by Dedukti. This requires handling tactics based on internal computation (reflective tactics), that we used intensively in our Coq proof. However, handling reflective tactics is currently not well supported by Dedukti. This is our current point of attention.

6.27. Performance improvements for a reflective tactic in the GeoCoq library

Participants: Pierre Boutry, Benjamin Grégoire, Enrico Tassi.

The GeoCoq library relies on a reflective tactic. It is an interesting topic to understand how to make such a tactic more efficient. A first pass on the algorithm makes that we manage to gain 15% of performance for the whole library and several orders of magnitude on specific subgoals. Another area of the tactic can also be improved by relying on Coq-Elpi.

6.28. Mutual interpretability of cartesian planes with Tarski's system of geometry

Participants: Pierre Boutry, Cyril Cohen.

A previous result by Pierre Boutry is that cartesian planes over pythagorean ordered fields are mutually interpretable with Tarski's system of geometry without the continuity axiom. This result can be extended by linking cartesian planes over real closed fields and the full Tarski system of geometry, understanding the continuity axiom as an implementation of Dedekind cuts. On the one hand, this requires a new proof that is not already found in the literature, on the other hand, this will result in a verified quantifier elimination procedure for Tarski's system of geometry, thus extending previous work by Cyril Cohen.

6.29. Simplification of a constructive version of Tarski's system of geometry

Participant: Pierre Boutry.

Our long term project is to show the independence of all thirteen axioms in a variant of Tarski's system of geometry. In the current situation, ten axioms have been checked to be independent using counter-models. Specific questions arise around the continuity axiom and decidability of equality between points. This is related to investigations concerning mutual interpretability with cartesian planes and an alternative system proposed by Michael Beeson.

6.30. Formal proofs of Tarjan's strongly connected components algorithm

Participants: Cyril Cohen, Laurent Théry, Ran Chen [Institute of Software, Chinese Academy of Science, Beijing], Jean-Jacques Lévy [Inria Paris, $\pi.r^2$ project-team], Stephan Merz [Inria Nancy Grand Est, Veridis project-team].

Comparing provers on a formalization of the same problem is always a valuable exercise. In this work, we present the formal proof of correctness of a non-trivial algorithm from graph theory that was carried out in three proof assistants: Why3, Coq, and Isabelle. This was published in an international conference [8].

SUMO Project-Team

7. New Results

7.1. New results on Axis 1: Quantitative models

7.1.1. Verification of Real-Time Models

Participants : Ocan Sankur, Nicolas Markey, Victor Roussanaly

7.1.1.1. Abstraction-refinement algorithms for model checking of timed automata.

The abstraction domain we consider [26] abstracts away zones by restricting the set of clock constraints that can be used to define them, while the refinement procedure computes the set of constraints that must be taken into consideration in the abstraction so as to exclude a given spurious counterexample. We implement this idea in two ways: an enumerative algorithm where a lazy abstraction approach is adopted, meaning that possibly different abstract domains are assigned to each exploration node; and a symbolic algorithm where the abstract transition system is encoded with Boolean formulas.

7.1.1.2. Robust controller synthesis problem in Büchi timed automata

We solve a robust controller synthesis problem [20] in a purely symbolic way. The goal of the controller is to play according to an accepting lasso of the automaton, while resisting to timing perturbations chosen by a competing environment. The problem was previously shown to be *PSPACE*-complete using regions-based techniques, but we provide a first tool solving the problem using zones only, thus more resilient to state-space explosion problem. The key ingredient is the introduction of branching constraint graphs allowing to decide in polynomial time whether a given lasso is robust, and even compute the largest admissible perturbation if it is. We also make an original use of constraint graphs in this context in order to test the inclusion of timed reachability relations, crucial for the termination criterion of our algorithm. Our techniques are illustrated using a case study on the regulation of a train network.

7.1.2. Verification of Stochastic Models

Participants : Hugo Bazille, Nathalie Bertrand, Éric Fabre, Blaise Genest, Ocan Sankur

7.1.2.1. Long-run satisfaction of path properties

We introduced the concepts of long-run frequency of path properties for paths in Kripke structures, and their generalization to long-run probabilities for schedulers in Markov decision processes [13]. We then studied the natural optimization problem of computing the optimal values of these measures, when ranging over all paths or all schedulers, and the corresponding decision problem when given a threshold. The main results are as follows. For (repeated) reachability and other simple properties, optimal long-run probabilities and corresponding optimal memoryless schedulers are computable in polynomial time. When it comes to constrained reachability properties, memoryless schedulers are no longer sufficient, even in the non-probabilistic setting. Nevertheless, optimal long-run probabilities for constrained reachability are computable in pseudo-polynomial time in the probabilistic setting and in polynomial time for Kripke structures. Finally for co-safety properties expressed by NFA, we gave an exponential-time algorithm to compute the optimal long-run frequency, and proved the PSPACE-completeness of the threshold problem.

7.1.2.2. Approximate Verification of Dynamic Bayesian Networks.

We are interested in studying the evolution of large homogeneous populations of cells, where each cell is assumed to be composed of a group of biological players (species) whose dynamics is governed by a complex biological pathway, identical for all cells. Modeling the inherent variability of the species concentrations in different cells is crucial to understand the dynamics of the population. In [9], we focus on handling this variability by modeling each species by a random variable that evolves over time. This appealing approach runs into the curse of dimensionality since exactly representing a joint probability distribution involving a large set of random variables quickly becomes intractable as the number of variables grows. To make this approach amenable to biopathways, we explore different techniques to (i) approximate the exact joint distribution at a given time point, and (ii) to track its evolution as time elapses.

7.1.2.3. Classification among stochastic systems

An important task in AI is one of classifying an observation as belonging to one class among several (e.g. image classification). We revisit this problem in a verification context: given k partially observable systems modeled as Hidden Markov Models (HMMs, also called labeled Markov chains), and an execution of one of them, can we eventually classify which system performed this execution, just by looking at its observations? Interestingly, this problem generalizes several problems in verification and control, such as fault diagnosis and opacity. Also, classification has strong connections with different notions of distances between stochastic models.

In [12], we study a general and practical notion of classifiers, namely limit-sure classifiers, which allow misclassification, i.e. errors in classification, as long as the probability of misclassification tends to 0 as the length of the observation grows. To study the complexity of several notions of classification, we develop techniques based on a simple but powerful notion of stationary distributions for HMMs. We prove that one cannot classify among HMMs iff there is a finite separating word from their stationary distributions. This provides a direct proof that classifiability can be checked in PTIME, as an alternative to existing proofs using separating events (i.e. sets of infinite separating words) for the total variation distance. Our approach also allows us to introduce and tackle new notions of classifiability which are applicable in a security context.

7.1.2.4. Fault diagnosis for stochastic systems

Diagnosis of partially observable stochastic systems prone to faults was introduced in the late nineties. Diagnosability, *i.e.* the existence of a diagnoser, may be specified in different ways: exact diagnosability requires that almost surely a fault is detected and that no fault is erroneously claimed; approximate diagnosability tolerates a small error probability when claiming a fault; last, accurate approximate diagnosability guarantees that the error probability can be chosen arbitrarily small.

In the article [7], we first refine the specification of diagnosability by identifying three criteria: (1) detecting faulty runs or providing information for all runs (2) considering finite or infinite runs, and (3) requiring or not a uniform detection delay. We then give a complete picture of relations between the different diagnosability specifications for probabilistic systems and establish characterisations for most of them in the finite-state case. Based on these characterisations, we develop decision procedures, study their complexity and prove their optimality. We also design synthesis algorithms to construct diagnosers and we analyse their memory requirements. Finally we establish undecidability of the diagnosability problems for which we provided no characterisation.

7.1.3. Energy Games

Participants : Loïc Hélouët, Nicolas Markey

7.1.3.1. Games with reachability objectives under energy constraints.

Under strict energy constraints (either only lower-bound constraint or interval constraint), we prove [23] that games with reachability objectives are LOGSPACE-equivalent to energy games with the same energy constraints but without reachability objective (i.e., for infinite runs). We then consider two kinds of relaxations of the upper-bound constraints (while keeping the lower-bound constraint strict): in the first one, called weak upper bound, the upper bound is absorbing, in the sense that it allows receiving more energy when the upper bound is already reached, but the extra energy will not be stored; in the second one, we allow for temporary violations of the upper bound, imposing limits on the number or on the amount of violations. We prove that when considering weak upper bound, reachability objectives require memory, but can still be solved in polynomial-time for one-player arenas; we prove that they are in co-NP in the two-player setting. Allowing for bounded violations makes the problem PSPACE-complete for one-player arenas and EXPTIME-complete for two players.

7.2. New results on Axis 2: Large Systems Models

7.2.1. Smart Transportation Systems

Participants : Nathalie Bertrand, Loïc Hélouët, Ocan Sankur

7.2.1.1. Smart regulation of urban train systems.

We have considered application of model checking techniques to evaluate performances of urban train systems [15]. Metros are subject to unexpected delays due to weather conditions, incidents, passenger misconduct, etc. To recover from delays and avoid their propagation to the whole network, metro operators use regulation algorithms that adapt speeds and departure dates of trains. Regulation algorithms are ad-hoc tools tuned to cope with characteristics of tracks, rolling stock, and passengers habits. However, there is no universal optimal regulation adapted in any environment. So, performance of a regulation must be evaluated before its integration in a network. In this work, we use probabilistic model-checking to evaluate the performance of regulation algorithms in simple metro lines. We model the moves of trains and random delays with Markov decision processes, and regulation as a controller that forces a decision depending on its partial knowledge of the state of the system. We then use the probabilistic model checker PRISM to evaluate performance of regulation: We compute the probability to reach a stable situation from an unstable one in less than d time units, letting d vary in a large enough time interval. This approach is applied on a case study, the metro network of Glasgow.

7.2.2. Supervisory Control

Participants : Hervé Marchand

7.2.2.1. Towards resilient supervisors against sensor deception attacks.

As a security problem, we considered in [24] feedback control systems where sensor readings may be compromised by a malicious attacker intent on causing damage to the system. We study this problem at the supervisory layer of the control system, using discrete event systems techniques. We assume that the attacker can edit the outputs from the sensors of the system before they reach the supervisory controller. In this context, we formulate the problem of synthesizing a supervisor that is robust against a large class of edit attacks on the sensor readings. The solution methodology is based on the solution of a partially observed supervisory control problem with arbitrary control patterns.

7.2.3. Multi-agent systems

Participants : Arthur Queffelec, Nicolas Markey, Ocan Sankur

7.2.3.1. Multi-agent path planning problems.

We are motivated by the increasing appeal of robots in information-gathering missions. In the problems we study [21], [22], the agents must remain interconnected. We model an area by a topological graph specifying the movement and the connectivity constraints of the agents. We study the theoretical complexity of the reachability and the coverage problems of a fleet of connected agents on various classes of topological graphs. We establish the complexity of these problems on known classes, and introduce a new class called sight-moveable graphs which admit efficient algorithms.

7.2.3.2. Quantitative semantics for Strategy Logic

We introduce and study SL[F], a quantitative extension of SL (Strategy Logic) [19], one of the most natural and expressive logics describing strategic behaviours. The satisfaction value of an SL[F] formula is a real value in [0,1], reflecting "how much" or "how well" the strategic on-going objectives of the underlying agents are satisfied. We demonstrate the applications of SL[F] in quantitative reasoning about multi-agent systems, by showing how it can express concepts of stability in multi-agent systems, and how it generalises some fuzzy temporal logics. We also provide a model-checking algorithm for our logic, based on a quantitative extension of Quantified CTL.

7.3. New results on Axis 3: Population Models

7.3.1. Verification

Participants : Nathalie Bertrand, Anirban Majumdar

7.3.1.1. Networks of many identical agents communicating by broadcast.

Broadcast networks allow one to model networks of identical nodes communicating through message broadcasts [17]. Their parameterized verification aims at proving a property holds for any number of nodes, under any communication topology, and on all possible executions. We focus on the coverability problem which dually asks whether there exists an execution that visits a configuration exhibiting some given state of the broadcast protocol. Coverability is known to be undecidable for static networks, i.e. when the number of nodes and communication topology is fixed along executions. In contrast, it is decidable in PTIME when the communication topology may change arbitrarily along executions, that is for reconfigurable networks. Surprisingly, no lower nor upper bounds on the minimal number of nodes, or the minimal length of covering execution in reconfigurable networks, appear in the literature. We showed tight bounds for cutoff and length, which happen to be linear and quadratic, respectively, in the number of states of the protocol. We also introduced an intermediary model with static communication topology and non-deterministic message losses upon sending. We showed that the same tight bounds apply to lossy networks, although, reconfigurable executions may be linearly more succinct than lossy executions. Finally, we showed NP-completeness for the natural optimisation problem associated with the cutoff.

7.3.1.2. Randomized distributed algorithms for consensus.

Randomized fault-tolerant distributed algorithms pose a number of challenges for automated verification: (i) parameterization in the number of processes and faults, (ii) randomized choices and probabilistic properties, and (iii) an unbounded number of asynchronous rounds. This combination makes verification hard. Challenge (i) was recently addressed in the framework of threshold automata. We extended threshold automata to model randomized consensus algorithms that perform an unbounded number of asynchronous rounds. For non-probabilistic properties, we showed [18] that it is necessary and sufficient to verify these properties under round-rigid schedules, that is, schedules where processes enter round r only after all processes finished round $r - 1$. For almost-sure termination, we analyzed these algorithms under round-rigid adversaries, that is, fair adversaries that only generate round-rigid schedules. This allowed us to do compositional and inductive reasoning that reduces verification of the asynchronous multi-round algorithms to model checking of a one-round threshold automaton. We applied this framework and automatically verified the following classic algorithms: Ben-Or's and Bracha's seminal consensus algorithms for crashes and Byzantine faults, 2-set agreement for crash faults, and RS-Bosco for the Byzantine case.

7.3.2. Control

Participants : Nathalie Bertrand, Blaise Genest, Anirban Majumdar

7.3.2.1. Controlling a population

We introduced a new setting where a population of agents [6], each modelled by a finite-state system, are controlled uniformly: the controller applies the same action to every agent. The framework is largely inspired by the control of a biological system, namely a population of yeasts, where the controller may only change the environment common to all cells. We studied a synchronisation problem for such populations: no matter how individual agents react to the actions of the controller, the controller aims at driving all agents synchronously to a target state. The agents are naturally represented by a non-deterministic finite state automaton (NFA), the same for every agent, and the whole system is encoded as a 2-player game. The first player (Controller) chooses actions, and the second player (Agents) resolves non-determinism for each agent. The game with m agents is called the m-population game. This gives rise to a parameterized control problem (where control refers to 2 player games), namely the population control problem: can Controller control the m-population game for all m $\in \mathbb{N}$ whatever Agents does? In this work, we proved that the population control problem is decidable, and it is a EXPTIME-complete problem. As far as we know, this is one of the first results on the control of parameterized systems. Our algorithm, which is not based on cut-off techniques, produces winning strategies which are symbolic, that is, they do not need to count precisely how the population is spread between states. The winning strategies produced by our algorithm are optimal with respect to the synchronisation time: the maximal number of steps before synchronisation of all agents in the target state is at most polynomial in the number of agents m, and exponential in the size of the NFA. We also showed that if there is no winning strategy, then there is a population size M such that Controller wins the m-population game if and only if m $\leq$ M. Surprisingly, M can be doubly exponential in the number of states of the NFA, with tight upper and lower bounds.

7.3.2.2. Concurrent multiplayer games with arbitrary many players

Traditional concurrent games on graphs involve a fixed number of players, who take decisions simultaneously, determining the next state of the game. In [16], we introduced a parameterized variant of concurrent games on graphs, where the parameter is precisely the number of players. Parameterized concurrent games are described by finite graphs, in which the transitions bear regular languages to describe the possible move combinations that lead from one vertex to another. We considered the problem of determining whether the first player, say Eve, has a strategy to ensure a reachability objective against any strategy profile of her opponents as a coalition. In particular Eve's strategy should be independent of the number of opponents she actually has. Technically, we focused on an *a priori* simpler setting where the languages labeling transitions only constrain the number of opponents (but not their precise action choices). These constraints are described as semilinear sets, finite unions of intervals, or intervals. We established the precise complexities of the parameterized reachability game problem, ranging from PTIME-complete to PSPACE-complete, in a variety of situations depending on the contraints (semilinear predicates, unions of intervals, or intervals) and on the presence or not of non-determinism.

7.4. New results on Axis 4: Data-driven Models

7.4.1. Crowdsourcing

Participants : Loïc Hélouët, Rituraj Singh

7.4.1.1. Complex workflows for crowdsourcing.

Crowdsourcing consists in hiring workers on internet to perform large amounts of simple, independent and replicated work units. We have proposed [32] complex workflows, a model for concurrent orcestration of tasks to solve problems that are more intricate than simpe tagging problems. Complex workflows allow higher-order answers where workers can suggest a process to obtain data rather than a plain answer. It is a data-centric model based on orchestration of concurrent tasks and higher order schemes. We have considered formal properties of specifications described with this model termination (whether some/all runs of a complex workflow terminate) and correctness (whether some/all runs of a workflow terminate with data satisfying FO requirements). We have shown that existential termination/correctness are undecidable in general excepted for specifications with bounded recursion. However, universal termination/correctness are decidable when constraints on inputs are specified in a decidable fragment of FO, and are at least in 2EXPTIME.

7.4.1.2. CrowdInc : a solution to reduce the cost of Consensus in Crowdsourcing.

Another contribution around crowdsourcing [34] considers agregation of answers, reliability of computed results, and optimization of costs. Crowdsoucing call for human expertise to solve problems which are still hard for computers, but easy for human workers. Crowdsourcing platform distribute replicated tasks to workers, pay them for their contribution, and aggregate answers to produce a reliable conclusion. A fundamental problem is to infer a correct answer from the set of results returned by workers. An additional ingredient of crowdsourcing is the cost needed to obtain a reliable answer: unlimited budget allows for the use of large pools of workers for each task, or experts to improve reliability of aggregated answers, but a limited budget forces to use resources at best to synthesize an reasonably reliable answer. We have focused on crowdsourcing of simple tasks with boolean answers. In this setting, we have first defined a probabilistic inference technique to agregate answers. This allows to consider difficulty of tasks and expertise of workers when building a conclusion. We have then proposed a greedy algorithm that reduces the cost (i.e. the number of workers hired by a platform) needed to reach a consensual answer. This algorithm considers difficulty of task, budget provided by client and total tasks to dynamically adapt threshold at each stage and makes locally optimal choice while preserving accuracy. Last, we have shown efficiency of our algorithm on several benchmarks, and compared its performance to existing solutions.

7.4.2. Guarded Attribute Grammars and Petri net synthesis

Participants : Adrian Puerto Aubel, Éric Badouel

7.4.2.1. Service-oriented programming

We addressed [30] the problem of component reuse for the design of user-centric distributed collaborative systems modelled by Guarded Attribute Grammars. Following the contract-based specification of components we develop an approach to an interface theory for the components of 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.

The visit of Joskel Ngoufo, a doctoral student at Yaoundé University, was the occasion to initiate a new implementation of the Guarded Attribute Grammars engine, in Racket language, a dialect of Lisp that allows metalanguage facilities and graphical interfaces to be processed more easily than in Haskell, the language chosen for the previous implementation.

7.4.2.2. Coordination of public debate.

Our research on data-centric collaborative systems has focused this year on the modelling of debates [28], with the aim of producing a tool that makes it possible to automatically conduct them, while managing relevant documents and analysing the respective positions of the different interventions from the point of view of argumentation theory. To this end, we are collaborating with Carlo Ferigato, a researcher at the JRC (C.E. Ispra, Italy), an institute for which we jointly produced a report covering an overview of the different theories developed around the subject, as well as the main tools proposing solutions to this problem. The aim of this collaboration is at understanding the basic principles and the computer programs apt to coordinate a public debate with an overall aim at giving the bases for designing such programs. Computer programs for the coordination of public debate exist since the beginning of the eighties but recently they have acquired new relevance for the use made of them by public administrations, associations and political parties. The meet of both citizen's needs and public administrations for transparency can today be technically realized with such programs through the present communication means in a more efficient way with respect to the first experiments dating now about forty years. This report aims at covering historical, technical and some theoretical aspects of the use of computers for the coordination of public debate.

7.4.2.3. Orthomodular partial orders.

The collaboration with Carlo Ferigato, is in line with the latter's thesis subject [11]. The set of regions of a condition/event transition system represents all the possible local states of a net system the behaviour of which is specified by the transition system. This set can be endowed with a structure, so as to form an orthomodular partial order. Given such a structure, one can then define another condition/event transition system. We study cases in which this second transition system has the same collection of regions as the first one. When it is so, the structure of regions is called stable. We proposed, to this aim, a composition operation, and a refinement operation for stable orthomodular partial orders, the results of which are stable.

7.5. New results on Transversal Concern: Missing Models

Participants : Hugo Bazille, Sihem Cherrared, Éric Fabre, Blaise Genest, Thierry Jéron, The Anh Pham

7.5.1. Unfolding-based dynamic partial-order reduction of asynchronous distributed programs

Unfolding-based Dynamic Partial Order Reduction (UDPOR) is a recent technique mixing Dynamic Partial Order Reduction (DPOR) with concepts of concurrency such as unfoldings to efficiently mitigate state space explosion in model-checking of concurrent programs. It is optimal in the sense that each Mazurkiewicz trace, *i.e.* a class of interleavings equivalent by commuting independent actions, is explored exactly once. In this work [25] we show that UDPOR can be extended to verify asynchronous distributed applications, where processes both communicate by messages and synchronize on shared resources. To do so, a general model of asynchronous distributed programs is formalized in TLA+. This allows to define an independence relation, a main ingredient of the unfolding semantics used by UDPOR during the UDPOR exploration. Then, the adaptation of UDPOR, involving the construction of an unfolding during the execution of the applicaton (*i.e.* with no model of the application but the code iteself), is made efficient by a precise analysis of dependencies. A prototype implementation gives promising experimental results.

7.5.2. Learning models for telecommunication management.

Model based methods have been recognised as the most appropriate approach to fault diagnosis in telecommunication networks, as they not only help in detecting and classifying failures, but is also provides useful explanations about the propagation of faults in such large distributed and concurrent systems. However, the bottleneck of these methods is of course the derivation and validation of a relevant model [8]. We have explored two techniques in this direction, based on fault/stress injection.

A first approach (collaboration Orange Labs) [33] consists in assembling generic components that would match the current (changing) topology of a software defined network. The model can then be validated by fault injection on a platform running the true VNF (virtual network functions) chains that are used in production. The second approach (collaboration Nokia Bell Labs) aims at detecting soft performance degradations, that would impact the quality of service, but not produce faults and alarms. Again, this can be achieved by stress injection at the level of VMs (virtual machines) in production software, and by collecting signature patterns under the form of statistical changes in the performance metrics collected on such systems.

7.5.3. Verification of deep neural networks.

Deep neural networks are as effective in their respective tasks as hardly understandable by a human. To use them in critical applications, not only they should be understood, they must be certified. We surveyed in [14] a large number of recent attempts to formally certify deep neural networks obtained by deep machine learning techniques. Most of the work currently focus on forward-propagating networks, and the problem of certifying their robustness.

TOCCATA Project-Team

7. New Results

7.1. Foundations and Spreading of Deductive Program Verification

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-Helft 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 [33].

GOSPEL - Providing OCaml with a Formal Specification Language In the context of the ANR project "VOCaL" (see Sec. 9.2.2), which aims at building a formally verified OCaml library of data structures and algorithms, a specification language for OCaml is designed and implemented. It is called GOSPEL, for Generic Ocaml SPEcification Language. During his post-doc in the Toccata team, from September 2018 to August 2019, C. Lourenço implemented a parser and type checker for GOSPEL. The work on the GOSPEL language has been presented at FM'19 [23]. J.-C. Filliâtre was keynote speaker at iFM 2019 and he gave a talk on the on-going work in the VOCaL project, including GOSPEL [17].

Program Verification Competition VerifyThis 2018 VerifyThis 2018 took place on April 14 and 15, 2018 in Thessaloniki, Greece, as a satellite event of ETAPS 2018. It was the sixth edition in the VerifyThis annual competition series. Typical challenges in the VerifyThis competitions are small but intricate algorithms given in pseudo-code with an informal specification in natural language. Participants have to formalize the requirements, implement a solution, and formally verify the implementation for adherence to the specification. There are no restrictions on the programming language and verification technology used. The time frame to solve each challenge is limited to 90 minutes. Submissions are judged for correctness, completeness, and elegance. The focus includes the usability of the tools, their facilities for formalizing the properties and providing helpful output.

VerifyThis 2018 consisted of three increasingly difficult verification challenges, selected to showcase various aspects of software verification. Eleven teams (one or two participants) took part in the competition. A full report on the VerifyThis 2018 event [39] provides a presentation of the competing teams, a description of the challenges, a high-level overview of the solutions, and the results of the competition.

Proof automation with the Coq proof assistant Proof assistants based on Type Theory, such as Coq, allow the implementation of effective automatic tactics based on computational reflection. These are usually limited to a particular mathematical domain (such as linear arithmetic or ring theory). In contrast, SMTCoq is a modular and extensible tool, using external provers, which generalizes these computational approaches to combine the reasoning from multiple domains. For this, it is based on a high-level interface, which offers greater expressiveness, at the cost of more complex automation. Q. Garchery and his co-authors [22] focused on two improvements: the ability to use quantified lemmas, and the ability to use multiple representations of the same data structure. They realized a new automatic tactic, based on SMTCoq, that is expressive while keeping the modularity and the efficiency of the latter. Such a tactic thus enable scalable, low-cost automation of new domains supported by state-of-the-art automatic provers.

Certificates for Logic Transformations In a context of formal program verification, using automatic provers, the trusted code base of verification environments is typically very broad. An environment such as Why3 implements many complex procedures: generation of verification conditions, logical transformations of proof tasks, and interactions with external provers. Considering only the logical transformations of Why3, their implementation already amounts to more than 17,000 lines of OCaml code. In order to increase our confidence in the correction of such a verification tool, Garchery, Keller, Marché and Paskevich present [32] proposed a mechanism of certifying transformations, producing certificates that can be validated by an external tool, according to the *skeptical* approach. They explored two methods to validate certificates: one based on a dedicated verifier developed in OCaml, the other based on the universal proof checker Dedukti. A specificity of their certificates is to be "small grains" and composable, which makes the approach incremental, allowing to gradually add new certifying transformations.

Reasoning About Universal Cubes in MCMT The Model Checking Modulo Theories (MCMT) framework is a powerful model checking technique for verifying safety properties of parameterized transition systems. In MCMT, logical formulas are used to represent both transitions and sets of states and safety properties are verified by an SMT-based backward reachability analysis. To be fully automated, the class of formulas handled in MCMT is restricted to cubes, i.e. existentially quantified conjunction of literals. While being very expressive, cubes cannot define properties with a global termination condition, usually described by a universally quantified formula. In this work, S. Conchon and M. Roux describe BRWP, an extension of the backward reachability of MCMT for reasoning about validity properties expressed as universal cubes, that is formulas of the form $\exists i \forall j. C(i, j)$, where $C(i, j)$ is a conjunction of literals. Their approach consists in a tight cooperation between the backward reachability loop and a deductive verification engine based on weakest-precondition calculus (WP). To provide evidence for the applicability of this new algorithm, they show how to make the model checker Cubicle cooperate with Why3 [25].

A Generalized Program Verification Workflow Based on Loop Elimination and SA Form. C. Lourenço, together with Maria Frade and Jorge Sousa Pinto from Universidade do Minho, developed a minimal model of the functioning of program verification and property checking tools based on (i) the encoding of loops as non-iterating programs, either conservatively, making use of invariants and assume/assert commands, or in a bounded way; and (ii) the use of an intermediate single-assignment (SA) form. The model captures the basic workflow of tools like Boogie, Why3, or CBMC, building on a clear distinction between operational and axiomatic semantics. This allows one to consider separately the soundness of program annotation, loop encoding, translation into SA form, and verification condition (VC) generation, as well as appropriate notions of completeness for each of these processes. To the best of our knowledge, this is the first formalization of a bounded model checking of software technique, including soundness and completeness proofs using Hoare logic; they also give the first completeness proof of a deductive verification technique based on a conservative encoding of invariant-annotated loops with assume/assert in SA form, as well as the first soundness proof based on a program logic. [21]

7.2. Reasoning on mutable memory in program verification

Certified Symbolic Execution Engine using Ghost Code Symbolic execution amounts to representing sets of concrete program states as a logical formula relating the program variables, and interpreting sets of executions as a transformation of that formula. B. Becker and C. Marché formalised the correctness of a symbolic interpreter engine, expressed by an over-approximation property stating that symbolic execution covers all concrete executions, and an under-approximation property stating that no useless symbolic states are generated. This formalisation is tailored for automated verification, that is the automated discharge of verification conditions to SMT solvers. To achieve this level of automation, they appropriately annotated the code of the symbolic interpreter with an original use of both ghost data and ghost statements [20].

Ghost Monitors M. Clochard, C. Marché and A. Paskevich proposed 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. This new 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. 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. 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.

Ghost monitors can be used to specify and prove fine-grained properties about the *infinite behaviors* of target programs. Since this cannot be easily done using existing verification frameworks, this work introduces a dedicated language for ghost monitors, with an original construction to *catch* and handle divergent executions. The soundness of the underlying program logic is established using a particular flavor of transfinite games. This language and its soundness are formalized and mechanically checked. [24]

7.3. Verification of Computer Arithmetic

Formal Verification of a State-of-the-Art Integer Square Root Even though some programs only use integer operations, the best way to understand and verify them might be to view them as fixed-point arithmetic algorithm. This is the case of the function from the GMP library that computes the square root of a 64-bit integer. The C code is short but intricate, as it implements Newton's method and it relies on magic constants and intentional arithmetic overflows. G. Melquiond and R. Rieu-Helft have verified this algorithm using the Why3 tool and automated solvers such as Gappa [28].

Round-off error and exceptional behavior analysis of explicit Runge-Kutta methods S. Boldo, F. Faissole, and A. Chapoutot developed a new fine-grained analysis of round-off errors in explicit Runge-Kutta integration methods, taking into account exceptional behaviors, such as underflow and overflow [12]. First steps towards the formalization has been done by F. Faissole [34].

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. D. Gallois-Wong, S. Boldo, and P. Cuoq proposed floating-point theorems that provide optimal bounds for all the intervals [13].

Emulating round-to-nearest-ties-to-zero "augmented" floating- point operations using round-to-nearest-ties-to-even ari
The 2019 version of the IEEE 754 Standard for Floating-Point Arithmetic recommends that new "augmented" operations should be provided for the binary formats. These operations use a new "rounding direction": round to nearest *ties-to-zero*. S. Boldo, C. Lauter, and J.-M. Muller show how they can be implemented using the currently available operations, using round-to-nearest *ties-to-even* with a partial formal proof of correctness [42].

LTI filters Several developments were made towards the efficency and accuracy of the implementation of LTI (linear time-invariant) numerical filters: a word-length optimization problem under accuracy constraints [26] by T. Hilaire, H. Ouzia, and B. Lopez, and a tight worst-case error analysis [16] by A. Volkova, T. Hilaire, and C. Lauter.

7.4. Spreading Formal Proofs

7.4.1. Real Analysis

Formally Verified Approximations of 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 [64]. 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 [14]. This makes CoqInterval one of the very few tools able to produce reliable results for improper integrals, be they formally verified or not.

Coq Formalization of algorithms for numerical filters D. Gallois-Wong developed a Coq formalization of a generic representation of numerical filters, called SIF [31] in order to encompass all other representations of filters, and prove useful theorems only once.

Complexity theory and constructive analysis E. Neumann and F. Steinberg extended the framework for complexity of operators in analysis devised by Kawamura and Cook (2012) to allow for the treatment of a wider class of representations and applied it to the study of interval computation [15]. A. Kawamura, F. Steinberg, and H. Thies put forward a complexity class of type-two linear-time [27].

F. Steinberg, L. Théry, and H. Thies give a number of formal proofs of theorems from the field of computable analysis. Results include that the algebraic operations and the efficient limit operator on the reals are computable, that certain countably infinite products are isomorphic to spaces of functions, compatibility of the enumeration representation of subsets of natural numbers with the abstract definition of the space of open subsets of the natural numbers, and that continuous realizability implies sequential continuity [46] [29]. F. Steinberg and H. Thies formalized proofs about Baire spaces and the isomorphy of the concrete and abstract spaces of open sets [45].

7.4.2. Formal Analysis of Debian packages

Several new results were produced in the context of the CoLiS project for the formal analysis of Debian packages. A first important step is the version 2 of the design of the CoLiS language done by B. Becker, C. Marché and other co-authors [38], that includes a modified formal syntax, a extended formal semantics, together with the design of concrete and symbolic interpreters. Those interpreters are specified and implemented in Why3, proved correct (following the initial approach for the concrete interpreter published in 2018 [60] and the recent approach for symbolic interpretation mentioned above [20]), and finally extracted to OCaml code.

To make the extracted code effective, it must be linked together with a library that implements a solver for feature constraints [61], and also a library that formally specifies the behavior of basic UNIX utilities. The latter library is documented in details in a research report [40].

A third result is a large verification campaign running the CoLiS toolbox on all the packages of the current Debian distribution. The results of this campaign were reported in another article [41] that will be presented at TACAS conference in 2020. The most visible side effect of this experiment is the discovery of bugs: more than 150 bugs report have been filled against various Debian packages.

7.4.3. Miscellaneous

Functional Programming. J.-C. Filliâtre was invited speaker at JFLA 2019, as part of a session celebrating the 30 years of JFLA (a French-speaking national conference related to functional programming). He talked about 25 years of programming with OCaml [18]. At JFLA 2020, J.-C. Filliâtre will give a talk related to the elimination of non-tail calls [30].

Formal Verification of “ParcourSup” algorithms. In May–July 2019, Léo Andrès (M1 student at Paris Sud) did a three month internship on the verification of the first algorithm of Parcoursup using Why3. Most of the expected properties, taken from the public description of Parcoursup’s algorithms, have been verified. Léo Andrès’s report (in French), is available on-line [37]. In June-December 2019, Benedikt Becker worked on the verification of the Java source code of ParcourSup. The findings and lessons learnt are described in a report under preparation.

Formalizing loop-carried dependencies in Coq for high-level synthesis. F. Faissole, G. Constantinides, and D. Thomas developed Coq formalizations in order to improve high-level synthesis for FPGAs [44].

VERIDIS Project-Team

7. New Results

7.1. Automated and Interactive Theorem Proving

Participants: Jasmin Christian Blanchette, Martin Bromberger, Antoine Defourné, Daniel El Ouraoui, Alberto Fiori, Mathias Fleury, Pascal Fontaine, Stephan Merz, Hamid Rahkooy, Hans-Jörg Schurr, Sorin Stratulat, Thomas Sturm, Sophie Tourret, Marco Voigt, Uwe Waldmann, Christoph Weidenbach.

7.1.1. Combination of Satisfiability Procedures

Joint work with Christophe Ringeissen (Inria Nancy – Grand Est, Pesto) and Paula Chocron (Insikt Intelligence, 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 [59]. 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 [60] 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 and 2019, we have been improving the framework and unified both results. This was published in the Journal of Automated Reasoning in 2019 [19].

7.1.2. Quantifier Handling in SMT

Joint work with 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.

In 2019, we investigated machine learning techniques for predicting the usefulness of an instance in order to decrease the number of instances passed to the SMT solver. For this, we proposed a meaningful way to characterize the state of an SMT solver, to collect instantiation learning data, and to integrate a predictor in the core of a state-of-the-art SMT solver. This ultimately leads to more efficient SMT solving for quantified problems.

7.1.3. Higher-Order SMT

Joint work with Haniel Barbosa, Andrew Reynolds, Cesare Tinelli (Univ. of Iowa), and Clark Barrett (Stanford)

SMT solvers have throughout the years been able to cope with increasingly expressive formulas, from ground logics to full first-order logic (FOL). In contrast, the extension of SMT solvers to higher-order logic (HOL) was mostly unexplored. We proposed a pragmatic extension for SMT solvers to support HOL reasoning natively without compromising performance on FOL reasoning, thus leveraging the extensive research and implementation efforts dedicated to efficient SMT solving. We showed how to generalize data structures and the ground decision procedure to support partial applications and extensionality, as well as how to reconcile quantifier instantiation techniques with higher-order variables. We also discussed a separate approach for redesigning an SMT solver for higher-order logic from the ground up via new data structures and algorithms. We applied our pragmatic extension to the CVC4 SMT solver and discussed a redesign of the veriT SMT solver. Our evaluation showed that they are competitive with state-of-the-art HOL provers and often outperform the traditional encoding into FOL.

This result was published at CADE 2019 [27]. We are also currently investigating extending the CCFV algorithm to higher-order logic.

7.1.4. Proofs for SMT

We have previously developed a framework for processing formulas in automatic theorem provers, with generation of detailed proofs that can be checked by external tools, including skeptical proof assistants. 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. In 2019, the format of proof output was further improved, while also improving the reconstruction procedure in the proof assistant Isabelle/HOL. This allowed the tactic using SMT with proofs to be regularly suggested by Sledgehammer as the fastest method to automatically solve proof goals. This was the subject of a workshop publication [36].

7.1.5. Clause Learning from Simple Models

The goal of this research is to guide inferences in expressive logics via simple models. Intuitively, a model is simple if computations with respect to the model can be done in polynomial time. We have shown that for first-order logic, models built from ground literals are sufficient to guide resolution inferences between non-ground clauses [35]. We have also investigated the expressivity of model representation formalisms in general [41]. Model representation formalisms built on atoms with only linear variable occurrences have the finite model property. Hence, they cannot represent infinite models.

7.1.6. SPASS-SATT

We have further developed our CDCL(T) solver SPASS-SATT. It is the combination of our SAT solver SPASS-SAT with highly efficient theory solvers for linear arithmetic [31]. SPASS-SATT showed good performance at the SMT competition 2019 where it won the category on linear rational arithmetic and scored second on linear integer arithmetic. The winner of the linear integer arithmetic category was a portfolio solver including SPASS-SATT. Our main improvements are due to an advanced clause normal form translation, a close interaction between the theory solvers and the SAT solver SPASS-SAT, and and a new transformation turning unbounded integer problems into bounded integer problems.

7.1.7. Extension of a Highly Efficient Prover to λ-free Higher-Order Logic

Joint work with Simon Cruanes (Aesthetic Integration), Stephan Schulz (DHBW Stuttgart), and Petar Vukmirović (VU Amsterdam).

Superposition-based provers, such as E, SPASS, and Vampire, are among the most successful reasoning systems for first-order logic. They serve as backends in various frameworks, including software verifiers, automatic higher-order theorem provers, and one-click “hammers” in proof assistants. Decades of research have gone into refining calculi, devising efficient data structures and algorithms, and developing heuristics to guide proof search. This work has mostly focused on first-order logic with equality, with or without arithmetic.

To obtain better performance, we propose to start with a competitive first-order prover and extend it to full higher-order logic one feature at a time. Our goal is a *graceful* extension, in keeping with the zero-overhead principle: *What you don't use, you don't pay for.*

As a stepping stone towards full higher-order logic, we initially restricted our focus to a higher-order logic without λ-expressions. Compared with first-order logic, its distinguishing features are partial application and applied variables. Our vehicle is E, a prover developed primarily by Schulz. It is written in C and offers good performance. E regularly scores among the top systems at the CASC competition, and usually is the strongest open source prover in the relevant divisions. It also serves as a backend for competitive higher-order provers.

Our experiments show that the λ-free higher-order version of E is practically as fast as E on first-order problems and can also prove higher-order problems that do not require synthesizing λ-terms. As a next step, we plan to add support for λ-terms and higher-order unification. This work is described in a TACAS 2019 conference paper [42]; an extended version of this paper has been invited to a special issue of the *International Journal on Software Tools for Technology Transfer*.

7.1.8. Extension of the Superposition Calculus with λ-Abstractions

Joint work with Alexander Bentkamp (VU Amsterdam) and Petar Vukmirović (VU Amsterdam).

We designed a superposition calculus for a clausal fragment of extensional polymorphic higher-order logic that includes anonymous functions but excludes Booleans. The inference rules work on $\beta\eta$-equivalence classes of λ-terms and rely on higher-order unification to achieve refutational completeness.

We implemented the calculus in the Zipperposition prover. Our empirical evaluation includes benchmarks from the TPTP (Thousands of Problems for Theorem Provers) and interactive verification problems exported from Isabelle/HOL. The results appear promising and suggest that an optimized implementation inside a competitive prover such as E, SPASS, or Vampire would outperform existing higher-order automatic provers. This research was presented at the CADE 2019 conference [28].

7.1.9. Automated Reasoning over Biological Networks

[54] study toricity of steady state ideals of biological models. From a computational point of view, models identified as toric allow to employ tools from toric geometry for a complexity reduction step. From a scientific point of view, toric models are known to have scale invariant multistationarity in the space of linear conserved quantities. This can be interpreted as a dimension reduction of the multistationarity problem. We propose a generalization of the notion of toricity, compatible with our above remarks, in terms of the geometry of the variety instead of the syntactic shape of generators of the ideal. We consider 129 models from the BioModels repository [67], for which ODEbase [0] provides input data directly usable for symbolic computation. While the existing literature was mostly limited to the complex numbers, we use real quantifier elimination methods to treat also the real case, which is clearly the relevant domain from a scientific point of view. In practice, our real computations in Redlog [4] can compete with our complex ones. In theory we show that our real algorithms are in EXPTIME while Gröbner bases, which are typically used when working with ideal generators, are EXPSPACE-complete [68]. To our knowledge, this is the first time that such a comprehensive set of biomodels has been systematically processed using symbolic methods.

7.1.10. Towards an Improved Encoding of TLA+ Proof Obligations

We reconsider the encoding of proof obligations that arise in proofs about TLA$^+$ specifications in multi-sorted first-order logic, and specifically their translations to SMT solvers. Our previous work [69] relied on type inference for identifying expressions having atomic types such as integers but did not exploit more complex types, even if such types were constructed during type inference. A more pervasive use of types for translating set-theoretic expressions to the input language of SMT solvers appears promising in order to reduce the use of type injections and quantifiers and thus simplify the proof obligations passed to the solver, but it raises non-trivial soundness and completeness issues. Techniques of gradual typing designed for programming languages where type inference is not fully possible statically may be helpful in this context. A related problem is support for instantiation hints for quantified formulas given by the user. A first paper will be presented at JFLA 2020.

[0] http://odebase.cs.uni-bonn.de/

7.1.11. Formal Proofs of Tarjan's Algorithm

Joint work with Ran Chen (Chinese Academy of Sciences), Cyril Cohen and Laurent Théry (Inria Sophia Antipolis Méditerranée, Stamp), and Jean-Jacques Lévy (Inria Paris, Pi.r2).

We consider Tarjan's classical algorithm for computing strongly connected components in a graph as a case study of intermediate complexity for comparing interactive proof assistants. Representing the algorithm as a functional program (rather than its more conventional imperative representation), we proved its correctness in three different proof assistants (Coq, Isabelle/HOL, and Why3). The proofs are based on essentially the same formulation of the algorithm and of its invariants, allowing us to compare differences due to idiosyncracies of the proof assistants, such as their ability to handle mutually recursive function definitions, proving termination beyond syntactic criteria, and their degree of automation. Our results were presented at ITP 2019 [33].

7.1.12. Implementation of an Efficient Validation of FOLID Cyclic Induction Reasoning

Checking the soundness of cyclic induction reasoning for first-order logic with inductive definitions ($\mathrm{FOL}_{\mathrm{ID}}$) is decidable but the standard checking method is based on an exponential complement operation for Büchi automata. We devised a polynomial method "semi-deciding" this problem; its most expensive steps are reminiscent of the comparisons with multiset path orderings. In practice, it has been integrated in the CYCLIST prover and successfully checked all the proofs included in its distribution. The work was presented at the CiSS2019 conference (Circularity in Syntax and Semantics) and the software is available at https://members.loria.fr/SStratulat/files/e-cyclist.zip.

7.2. Formal Methods for Developing and Analyzing Algorithms and Systems

Participants: Étienne André, Marie Duflot-Kremer, Yann Duplouy, Margaux Duroeulx, Igor Konnov, Dominique Méry, Stephan Merz, Nicolas Schnepf, Christoph Weidenbach.

7.2.1. Synthesis of Security Chains for Software Defined Networks

Joint work with Rémi Badonnel and Abdelkader Lahmadi (Inria Nancy – Grand Est, Resist).

The PhD thesis of Nicolas Schnepf focuses on applying techniques based on formal methods in the area of network communications, and in particular for the construction, verification, and optimization of chains of security functions in the setting of software-defined networks (SDN). The main objective is to prevent applications from disrupting the functioning of the network or services, for example by launching denial of service attacks, port scanning or similar activities.

We designed techniques for formally verifying security chains using SMT solving and symbolic model checking. Furthermore, we developed and prototypically implemented an approach for (i) learning a Markov chain characterizing the network behavior of an Android application based on its observed communications, (ii) inferring appropriate security functions from the structure of that Markov chain and thresholds set by the network operator, using techniques of logic programming, (iii) combining security functions for individual applications into larger security chains, and (iv) optimizing the deployment of security chains for a given SDN infrastructure using techniques of (linear or non-linear) optimization or optimizing SMT solvers. Two papers were presented at IM 2019 [39], [38], the PhD thesis [12] was defended in September 2019, and a journal paper is in preparation.

7.2.2. Satisfiability Techniques for Reliability Assessment

Joint work with Nicolae Brînzei at Centre de Recherche en Automatique de Nancy.

In the context of the PhD thesis of Margaux Durœulx, funded by the Lorraine University of Excellence program, we explore the applicability of satisfiability techniques for assessing the reliability of complex systems. In particular, we consider component-based systems modeled using fault trees that can be seen as a visual representation of the structure function indicating which combinations of component failures lead to system failures. We rely on SAT solvers to compute minimal tie sets, i.e., minimal sets of components whose functioning ensures that the overall system works. These tie sets are instrumental for a probabilistic reliability assessment. In 2019, we have extended this idea to dynamic fault trees where the order of component failures needs to be taken into account in order to determine the failure status of the overall system [34].

7.2.3. Statistical Model Checking of Distributed Programs

Yann Duplouy joined the HAC SPECIS project (cf. section 9.2) in December 2018 as a post-doctoral researcher with the objective of designing and implementing a statistical model checker within the SimGrid framework. So far he added to SimGrid the possibility to use stochastic profiles, introducing probabilities in the model of the network. He also developed a prototype tool that can be interfaced with the SimGrid simulators to perform statistical model checking on the actual programs simulated using the SimGrid framework. He now validates this prototype on concrete case studies, including the Bit Torrent protocol with probabilistic failures of the nodes.

7.2.4. Parameterized Verification of Threshold-Guarded Fault-Tolerant Distributed Algorithms

Joint work with Nathalie Bertrand (Inria Rennes Bretagne – Atlantique, SUMO), Marijana Lazić (TU Munich) and 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 [62], [63], [64].

Our previous results apply to asynchronous algorithms. It is well-known that distributed consensus cannot be solved in purely asynchronous systems [61]. However, when an algorithm is provided with a random coin, consensus becomes solvable (e.g., the algorithm by Ben-Or, 1993). In [29], 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 and randomized one-step consensus (RS-BOSCO [71]). The verification of the latter algorithm required us to run experiments in Grid5000. This paper was presented at CONCUR 2019.

Another way of making consensus solvable is to impose synchrony on the executions of a distributed system. In [40] 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. This paper was presented at TACAS 2019. The paper was invited to the special issue of TACAS 2019, to appear in the *International Journal on Software Tools for Technology Transfer* in 2020.

7.2.5. 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 [66]. 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.

Following our results in 2018 [65], we have extended the symbolic model checker for TLA$^+$. In [22], we have defined the translation process from TLA$^+$ to SMT as a series of rewriting rules. Furthermore, we have proven soundness of this translation. Our experiments show that APALACHE runs faster than TLC when proving inductive invariants. APALACHE also implements bounded model checking, which has to be improved, in order to make it competitive with TLC. The paper [22] was presented at ACM OOPSLA 2019.

7.2.6. Incremental Development of Systems and Algorithms

Joint work with Rosemary Monahan (NUI Maynooth, Ireland) and Mohammed Mosbah (LaBRI, Bordeaux).

The development of distributed algorithms and, more generally, of distributed systems, is a complex, delicate, and challenging process. The approach based on refinement applies a design methodology that starts from the most abstract model and leads, in an incremental way, to a distributed solution. The use of a proof assistant gives a formal guarantee on the conformance of each refinement with the model preceding it. Our main result during 2019 is the development of a distributed pattern [26] handling the dynamicity of the topology of networks.

[0] WWTF project APALACHE (ICT15-103): https://forsyte.at/research/apalache/

CIDRE Project-Team

6. New Results

6.1. Axis 1 : Attack comprehension

6.1.1. Fault injection

Electromagnetic injection is a non-invasive way to attack a chip. The large number of parameters that require to be properly tuned for such an attack limits its efficiency. In [30] we propose several ways to improve the success rate of fault injection by electromagnetic radiation. We show that software execution is altered at targeted instructions if the radiating probe is located above the phase-locked loop device driving the clock tree. We identify the phase-locked loop as a sensitive part of the chip. We reduce the preferential location for the electromagnetic injection to a small area in the vicinity of the analog power supply feeding the phase-locked loop. We also explore the influence of the frequency of the injected electromagnetic wave. We compute the optimal fault rate in a bandwidth of $15MHz$, in the upper limit of the chip bandwidth. Our experiments show that for an optimal frequency a precision of $5ns$, we succeed to reach the best fault rate. With this electromagnetic injection technique, the achieved success rate reaches 15 to 20%. Such a fault can be used to retrieve the key of a cryptographic algorithm (for an Advanced Encryption Standard application for example).

6.1.2. Malware analysis

About Android malware analysis, we have started investigations with specific malware that hide their behavior using obfuscation techniques [10]. As these malware are difficult to find in the wild, we have also started to analyze both datasets of the literature and large collection of applications captured from different repositories such as the Play Store. This huge amount of applications to analyze (currently more than 100,000) makes difficult to build reliable experiments [20]. We have designed a new tool, called PyMaO, that helps to orchestrate experiments. This tool is published as an open source tool under GPL v3. We have also revisited the historical datasets of malware of the literature and introduce a more up-to-date malware and goodware dataset [26].

6.1.3. Focus on doxware

A doxware is a particular type of ransomware that threatens to release personal or sensitive data to the public if the user does not pay the ransom. The term comes from the hacker term "doxing," or releasing confidential information over the internet. The only difference between a classical ransomware and a doxware resides in a *valuable files hunting* followed by an exfiltration of these data. In [34], we have explored how an attacker may be able to quickly localized valuable assets of a machine using an analysis of the content and the vocabulary of its files.

6.1.4. Attack scenario reconstruction

In order to supervise the security of a large infrastructure, the administrator deploys multiple sensors and intrusion detection systems on several critical places in the system. It is easier to explain and detect attacks if more events are logged. Starting from a suspicious event (appearing as a log entry), the administrator can start his investigation by manually building the set of previous events that are linked to this event of interest. Accordingly, the administrator attempts to identify links among the logged events in order to retrieve those that correspond to the traces of the attacker's actions in the supervised system; previous work is aimed at building these connections. In practice, however, this type of link is not trivial to define and discover. Hence, there is a real necessity to describe and define formally the semantics of these links in literature. In In order to supervise the security of a large infrastructure, the administrator deploys multiple sensors and intrusion detection systems on several critical places in the system. It is easier to explain and detect attacks if more events are logged. Starting from a suspicious event (appearing as a log entry), the administrator can start his investigation by manually building the set of previous events that are linked to this event of interest. Accordingly, the administrator attempts to identify links among the logged events in order to retrieve those that correspond to the traces of the attacker's actions in the supervised system; previous work is aimed at building these connections. In practice, however, this type of link is not trivial to define and discover. Hence, there is a real necessity to describe and define formally the semantics of these links in literature. In this paper, a clear definition of this relationship, called contextual event causal dependency, is introduced and proposed. The work presented in this paper aims at defining a formal model that would ideally unify previous work on causal dependencies among heterogeneous events. We define a relationship among events that enables the discovery of all events, which can be considered as the cause (in the past) or the effect (in the future) of an event of interest (e.g., an indicator of compromise, produced by an attacker action). In [36], we have proposed a clear definition of this relationship, called contextual event causal dependency. The work presented in [36] aims at defining a formal model that would ideally unify previous work on causal dependencies among heterogeneous events. We define a relationship among events that enables the discovery of all events, which can be considered as the cause (in the past) or the effect (in the future) of an event of interest (e.g., an indicator of compromise, produced by an attacker action).

6.2. Axis 2 : Attack detection

6.2.1. Vulnerabilities detection in Java

In a prior work, we have focused on adapting a machine-learning tool (ChuckyJava) aiming at automatically detect vulnerabilities in Java. ChuckyJava is able to detect vulnerabilities by performing in two steps: the neighborhood discovery and the anomaly detection. The neighborhood discovery is the ability for the tool to detect method of similar semantics: neighbors. In [25], we mitigate many ChuckyJava's limitations by developing JavaNeighbors that improves the neighborhood discovery. JavaNeighbors represents methods by terms and using a method based on a Natural Language Processing technique, JavaNeighbors computes the distance between all representations of methods. Finally, according to the distance, each method has a neighbor list from the closest to the most distant ones. JavaNeighbors has enabled ChuckyJava to detect vulnerabilities with more accuracy.

6.2.2. Ransomware detection

A ransomware attacks mostly begins with social engineering methods to install payloads on victims' computers, followed by a communication with command and control servers for data exchange. To enable an early detection and thus scale down these attacks, we propose in [35] a detection model based on the collected system and network logs from a computer. The analysis is performed on various ransomware families with a high detection rate. Packet level detection is performed to grant the best use case scenario. This work intends to provide an independent third-party procedure that is able to distinguish between a benign software and a malicious ransomware based on network activity. Furthermore, it is not limited to only identify ransomware but could be utilized to inspect different malware.

6.2.3. Intrusion detection using logs of distributed application

Although security issues are now addressed during the development process of distributed applications, an attack may still affect the provided services or allow access to confidential data. To detect intrusions [22], we consider an anomaly detection mechanism which relies on a model of the monitored application's normal behavior. During a model construction phase, the application is run multiple times to observe some of its correct behaviors. Each gathered trace enables the identification of significant events and their causality relationships, without requiring the existence of a global clock. The constructed model is dual: an automaton plus a list of likely invariants. The redundancy between the two sub-models decreases when generalization techniques are applied on the automaton. Solutions already proposed suffer from scalability issues. In particular, the time needed to build the model is important and its size impacts the duration of the detection phase. The proposed solutions address these problems, while keeping a good accuracy during the detection phase, in terms of false positive and false negative rates. To evaluate them, a real distributed application and several attacks against the service have been considered. One of our goal is to identify redundancies and complementarities between the proposed models.

6.3. Axis 3 : Attack resistance

6.3.1. Attacker Life cycle

We have been witnessing for years the awareness of the existence of a so-called Advanced Persistent Threat (APT). These attacks, regularly target or involving nation-states and large companies, were first defined in 2011. Ad Advanced Persistent Threat: (i) pursues its objectives repeatedly over an extended period of time; (ii) adapts to defenders' efforts to resist it; and (iii) is determined to maintain the level of interaction needed to execute its objectives. In [13], we have proposed a model providing an operational reading of the attackers' lifecycle in a compromised network. This model allows to express possible regressions in the attack and introduces the concept of a waiting state, which is essential for long-term actions. In this article we have also proposed a confrontation between our model and two recent examples of attacks whose progression has been publicly described: the Equifax breach (2017) and the TV5Monde sabotage (2015).

6.3.2. OS-level intrusion survivability

Despite the deployment of preventive security mechanisms to protect the assets and computing platforms of users, intrusions eventually occur. In [17], we have proposed a novel intrusion survivability approach to withstand ongoing intrusions. Our approach relies on an orchestration of fine-grained recovery and per-service responses (e.g., privileges removal). Such an approach may put the system into a degraded mode. This degraded mode prevents attackers to reinfect the system or to achieve their goals if they managed to reinfect it. It maintains the availability of core functions while waiting for patches to be deployed. We devised a cost-sensitive response selection process to ensure that while the service is in a degraded mode, its core functions are still operating. We built a Linux-based prototype and evaluated the effectiveness of our approach against different types of intrusions. The results show that our solution removes the effects of the intrusions, that it can select appropriate responses, and that it allows services to survive when reinfected. In terms of performance overhead, in most cases, we observed a small overhead, except in the rare case of services that write many small files asynchronously in a burst, where we observed a higher but acceptable overhead.

6.3.3. Secure routing in drones swarms

Unmanned aerial vehicle (UAV) applications and development have increased over the past few years as this technology has become more accessible and less expensive. On a single UAV scenario, communication is a keystone to transmit commands and retrieve data from UAV sensors. It is even more critical in swarm where cooperation and inter messaging is fundamental. The communication between the nodes of a swarm is based on a suitable routing algorithm. The routing must allow each node to send messages to each other, by successive hops between different neighbors. A UAV swarm is a particular mobile ad-hoc networks where nodes run independently but form a cooperative communication network. UAV swarm shares com- mon characteristics with VANET (vehicular ad hoc network), sensors network or mobile phone network but also strongly differs on specific points (mobility model, instability, limited infrastructure access). Any computation on a UAV is a permanent trade off between volume, weight and power consumption, with no infrastructure access. In [31], we have proposed a secured routing protocol designed for UAV swarm networks. SEER4US is the first protocol providing integrity of routing messages and authentication of their sender with low energy consumption for battery preservation.

6.3.4. Securing the control flow of smartcard C programs

Results obtained several years ago about securing the control flow of C programs have been extended and published in the journal Computers and Security [7]. This extended version of our work focuses on the formal verification of the introduced countermeasures. We prove that any possible attack that would skip more than one C instruction is detected by our countermeasures. We also extended the experimental results on a benchmark software dedicated to smartcards. This work has been achieved in cooperation with Karine Heydemann from the LIP6 laboratory (Sorbonne Université).

6.3.5. A secure implementation of the replicated state machine

State machine replication (RSM) is today the foundation of many cloud-based highly-available products: it allows some service to be deployed such to guarantee its correct functioning despite possible faults. In RSM, clients issue operation requests to a set of distributed processes implementing the replicated service, that, in turn, run a protocol to decide the order of execution of incoming operations and provide clients with outputs. Faults can be accidental (e.g. a computer crashing due to a loss of power) or have a malicious intent (e.g. a compromised server). Whichever is the chosen fault model, RSM has proven to be a reliable and effective solution for the deployment of dependable services. RSM is usually built on top of a distributed Consensus primitive that is used by processes to agree on the order of execution of requests concurrently issued by clients. The main problem with this approach is that Consensus is impossible to achieve deterministically in a distributed settings if the system is asynchronous and even just a single process may fail by crashing. This led the research community to study and develop alternative solutions based on the relaxation of some of the constraints, to allow agreement to be reached in partially synchronous systems with faulty processes by trading off consistency with availability. An alternative approach consists in imposing constraints on the set of operations that can be issued by clients, i.e. imposing updates that commute. In particular, commutative replicated data types (CRDTs) can be implemented with an RSM approach in asynchronous settings using the monotonic growth of a join semilattice, i.e., a partially ordered set that defines a join (least upper bound) for all element pairs. In [18] we have proposed an algorithm that solves Generalized Lattice Agreement in a Byzantine fault model. To the best of our knowledge this is the first solution for Byzantine lattice agreement that works on any possible lattice, and it is the first work proposing a Byzantine tolerant RSM built on it. The algorithm is wait-free, i.e., every process completes its execution of the algorithm within a bounded number of steps, regardless of the execution of other processes. We have also sketch the main lines of a signature-based version of our algorithms which take advantage of digital signatures to reduce the message complexity to $\mathcal{O}(n)$ per process, when the number f of Byzantine processes verifies $f = \mathcal{O}(1)$.

6.3.6. Blockchain in adversarial environments

We are pursuing our efforts dedicated to the theoretical aspects of blockchains. In particular, we have recently proposed to specify blockchains as a composition of abstract data types all together with a hierarchy of consistency criteria that formally characterizes the histories admissible for distributed programs that use them. Our work is based on an original oracle-based construction that, along with new consistency definitions, captures the eventual convergence process in blockchain systems. This study allows us to focus on the implementability of the presented abstractions and a mapping of representative existing blockchains from both academia and industry in our framework. It is already known that some blockchain implementations solve eventual consistency of an append-only queue using Consensus. However the question about the consistency criterion of blockchains as Bitcoin and Ethereum that technically do not solve Consensus, and their relation with Consensus in general was not studied. We have also proposed a specification of distributed ledger register that matches the Lamport hierarchy from safe to atomic. Moreover, we propose implementations of distributed ledger registers with safe, regular and atomic guaranties in a model of communication specific to distributed ledgers technology that we also formalize. Then, we propose an implementation of a distributed ledger register that satisfies the atomic specification and the k-consistency property that characterizes the permissionless distributed blockchains such as Bitcoin and Ethereum. Preliminary results appear in [41].

In parallel to this work, we have proposed the design of a scalable permissionless blockchain in the proof-of-stake setting. In particular, we use a distributed hash table as a building block to set up randomized shards, and then leverage the sharded architecture to validate blocks in an efficient manner. We combine verifiable Byzantine agreements run by shards of stakeholders and a block validation protocol to guarantee that forks occur with negligible probability. We impose induced churn to make shards robust to eclipse attacks, and we rely on the UTXO coin model to guarantee that any stake-holder action is securely verifiable by anyone. Our protocol works against adaptive adversary, and makes no synchrony assumption beyond what is required for the byzantine agreement. This work has been published in [19].

COMETE Project-Team

7. New Results

7.1. Foundations of privacy and quantitative information flow

Privacy and information flow have the common goal of trying to protect sensitive information. Comete focuses in particular on the potential leaks due to inference from data that are public, or anyway available to the adversary. We consider the probabilistic aspects, and we use concepts and tools from information theory.

7.1.1. Black-box Leakage Estimation

In [16] we have considered the problem of measuring how much a system reveals about its secret inputs under the black-box setting. Black-box means that we assume no prior knowledge of the system's internals: the idea is to run the system for choices of secrets and measure its leakage from the respective outputs. Our goal was to estimate the Bayes risk, from which one can derive some of the most popular leakage measures (e.g., min-entropy, additive, and multiplicative leakage). The state-of-the-art method for estimating these leakage measures is the frequentist paradigm, which approximates the system's internals by looking at the frequencies of its inputs and outputs. Unfortunately, this does not scale for systems with large output spaces, where it would require too many input-output examples. Consequently, it also cannot be applied to systems with continuous outputs (e.g., time side channels, network traffic). In [16] we have exploited an analogy between Machine Learning (ML) and black-box leakage estimation to show that the Bayes risk of a system can be estimated by using a class of ML methods: the universally consistent learning rules; these rules can exploit patterns in the input-output examples to improve the estimates' convergence, while retaining formal optimality guarantees. We have focused on a set of them, the nearest neighbor rules; we show that they significantly reduce the number of black-box queries required for a precise estimation whenever nearby outputs tend to be produced by the same secret; furthermore, some of them can tackle systems with continuous outputs. We have illustrated the applicability of these techniques on both synthetic and real-world data, and we compared them with the state-of-the-art tool, leakiEst, which is based on the frequentist approach.

7.1.2. An Axiomatization of Information Flow Measures

Quantitative information flow aims to assess and control the leakage of sensitive information by computer systems. A key insight in this area is that no single leakage measure is appropriate in all operational scenarios; as a result, many leakage measures have been proposed, with many different properties. To clarify this complex situation, in [11] we have studied information leakage axiomatically, showing important dependencies among different axioms. We have also established a completeness result about the g-leakage family, showing that any leakage measure satisfying certain intuitively-reasonable properties can be expressed as a g-leakage.

7.1.3. Comparing systems: max-case refinement orders and application to differential privacy

Quantitative Information Flow (QIF) and Differential Privacy (DP) are both concerned with the protection of sensitive information, but they are rather different approaches. In particular, QIF considers the expected probability of a successful attack, while DP (in both its standard and local versions) is a max-case measure, in the sense that it is compromised by the existence of a possible attack, regardless of its probability. Comparing systems is a fundamental task in these areas: one wishes to guarantee that replacing a system A by a system B is a safe operation, that is the privacy of B is no-worse than that of A. In QIF, a refinement order provides strong such guarantees, while in DP mechanisms are typically compared (wrt privacy) based on the ε privacy parameter that they provide.

In [15] we have explored a variety of refinement orders, inspired by the one of QIF, providing precise guarantees for max-case leakage. We have studied simple structural ways of characterizing them, the relation between them, efficient methods for verifying them and their lattice properties. Moreover, we have applied these orders in the task of comparing DP mechanisms, raising the question of whether the order based on ε provides strong privacy guarantees. We have shown that, while it is often the case for mechanisms of the same "family" (geometric, randomised response, etc.), it rarely holds across different families.

7.1.4. A Logical Characterization of Differential Privacy

Differential privacy (DP) is a formal definition of privacy ensuring that sensitive information relative to individuals cannot be inferred by querying a database. In [12], we have exploited a modeling of this framework via labeled Markov Chains (LMCs) to provide a logical characterization of differential privacy: we have considered a probabilistic variant of the Hennessy-Milner logic and we have defined a syntactical distance on formulae in it measuring their syntactic disparities. Then, we have defined a trace distance on LMCs in terms of the syntactic distance between the sets of formulae satisfied by them. We have proved that such distance corresponds to the level of privacy of the LMCs. Moreover, we have used the distance on formulae to define a real-valued semantics for them, from which we have 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 have focused on bisimulation semantics on nondeterministic probabilistic processes and we have provided 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 have shown that the generalized bisimulation distance on processes is equal to the syntactic distance on their mimicking formulae. Moreover, we have used the distance on mimicking formulae to obtain bounds on differential privacy.

7.1.5. Geo-indistinguishability vs Utility in Mobility-based Geographic Datasets

In [17] we have explored the trade-offs between privacy and utility in mobility-based geographic datasets. Our aim was to find out whether it is possible to protect the privacy of the users in a dataset while, at the same time, maintaining intact the utility of the information that it contains. In particular, we have focused on geo-indistinguishability as a privacy-preserving sanitization methodology, and we have evaluated its effects on the utility of the Geolife dataset. We have tested the sanitized dataset in two real world scenarios: 1. Deploying an infrastructure of WiFi hotspots to offload the mobile traffic of users living, working, or commuting in a wide geographic area; 2. Simulating the spreading of a gossip-based epidemic as the outcome of a device-to-device communication protocol. We have shown the extent to which the current geo-indistinguishability techniques trade privacy for utility in real world applications and we focus on their effects at the levels of the population as a whole and of single individuals.

7.1.6. Utility-Preserving Privacy Mechanisms for Counting Queries

Differential privacy(DP) and local differential privacy(LPD) are frameworks to protect sensitive information in data collections. They are both based on obfuscation. In DP the noise is added to the result of queries on the dataset, whereas in LPD the noise is added directly on the individual records, before being collected. The main advantage of LPD with respect to DP is that it does not need to assume a trusted third party. 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 sanitized data it is necessary to have a huge collection of them. In [25], we focus on the problem of estimating counting queries from collections of noisy answers, and we propose a variant of LDP based on the addition of geometric noise. Our main result is that the geometric noise has a better statistical utility than other LPD mechanisms from the literature.

7.1.7. Differential Inference Testing: A Practical Approach to Evaluate Sanitizations of Datasets

In order to protect individuals' privacy, data have to be "well-sanitized" before sharing them, i.e. one has to remove any personal information before sharing data. However, it is not always clear when data shall be deemed well-sanitized. In this paper, we argue that the evaluation of sanitized data should be based on whether the data allows the inference of sensitive information that is specific to an individual, instead of being centered around the concept of re-identification. In [20] we have proposed a framework to evaluate the effectiveness of different sanitization techniques on a given dataset by measuring how much an individual's record from the sanitized dataset influences the inference of his/her own sensitive attribute. Our intent was not to accurately predict any sensitive attribute but rather to measure the impact of a single record on the inference of sensitive information. We have demonstrated our approach by sanitizing two real datasets in different privacy models and evaluate/compare each sanitized dataset in our framework.

7.2. Foundations of Process Calculi

7.2.1. Group Distributed Knowledge.

We introduced spatial constraint systems (scs) as semantic structures for reasoning about spatial and epistemic information in concurrent systems. They have been used to reason about beliefs, lies, and group epistemic behaviour inspired by social networks. They have also been used for proving new results about modal logics and giving semantics to process calculi. In [19] we developed the theory of scs to reason about the distributed information of potentially infinite groups. We characterized the notion of distributed information of a group of agents as the infimum of the set of join-preserving functions that represent the spaces of the agents in the group. We provided an alternative characterization of this notion as the greatest family of join-preserving functions that satisfy certain basic properties. We showed compositionality results for these characterizations and conditions under which information that can be obtained by an infinite group can also be obtained by a finite group. Finally, we provided algorithms that compute the distributive group information of finite groups. Furthermore, in [14] we summarized all the main results we have obtained about scs.

7.2.2. Group Polarization.

Social networks can make their users become more radical and isolated in their own ideological circle causing dangerous splits in society in a phenomenon known as group polarization. In [22] we developed a preliminary model for social networks, and a measure of the level of polarization in these social networks, based on Esteban and Ray's classic measure of polarization for economic situations. Our model includes information about each agent's quantitative strength of belief in a proposition of interest and a representation of the strength of each agent's influence on every other agent. We considered how the model changes over time as agents interact and communicate, and included several different options for belief update, including rational belief update and update taking into account irrational responses such as confirmation bias and the backfire effect. Under various scenarios, we considered the evolution of polarization over time, and the implications of these results for real world social networks.

7.2.3. Lattice Theory.

Structures involving a lattice and join-endomorphisms on it are ubiquitous in computer science. In [28] we studied the cardinality of the set $J(L)$ of all join-endomorphisms of a given finite lattice L. We showed that the cardinality of $J(L)$ is sub-exponential, exponential and super-exponential in the size of the lattice for boolean algebras, linear-orders, and arbitrary lattices, respectively. We also studied the following problem: Given a lattice L of size n and a subset S of $J(L)$ of size m, find the greatest lower bound in $J(L)$ of S. This join-endomorphism has meaningful interpretations in epistemic logic, distributed systems, and Aumann structures. We showed that this problem can be solved with worst-case time complexity in $O(n + mlogn)$ for powerset lattices, $O(mn^2)$ for lattices of sets, and $O(mn + n^3)$ for arbitrary lattices. The complexity is expressed in terms of the basic binary lattice operations performed by the algorithm.

7.2.4. Festschrift Contribution.

In a Festschrift dedicated to Catuscia Palamidessi [26], we presented an article with original solutions to four challenging mathematical puzzles [23]. The first two are concerned with random processes. The first problem can be reduced to computing, for arbitrary large values of n, the expected number of iterations of a program that increases a variable at random between 1 and n until exceeds n. The second problem can be reduced to determining the probability of reaching a given point after visiting all the others in a circular random walk. The other two problems involve finding optimal winning group strategies in guessing games.

DATASPHERE Team

6. New Results

6.1. 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 [5], we showed how the movement above ground of the intermediation activity transforms territories.

6.2. 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 [4]. 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 specificity of Asia in these transformations [6].

6.3. Network data analytics

In collaboration with the Chinese Academy of Sciences, we worked on packet processing algorithmic for high speed network measurements. In [1] a packet capture archive system is developed and described. 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 [3].

6.4. Geopolitics of BGP

We have investigated the logical layer of cyberspace through an analysis of the structure of connectivity and the Border Gateway Protocol (BGP). This protocol has been leveraged by countries to control the flow of information or for active strategic purposes. We focused on several countries and characterized their strategies by linking them to current architectures and understanding their resilience in times of crisis. We focus on the case of Iran and uncovered the deep changes that has happened in the past 5 years. This study was premonitory as we observed in november 2019 the full extend of the strategy with the large scale internet disruptions. This generates a lot of mediatic coverage.

PESTO Project-Team

7. New Results

7.1. Security protocols

7.1.1. Analysis of Equivalence Properties

Participants: Vincent Cheval, Véronique Cortier, Ivan Gazeau, Steve Kremer, 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 DEEPSEC tool is a recent 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). To improve efficiency for non-determinate processes, Cheval, Kremer and Rakotonirina [21] develop new optimisation techniques. This is achieved through a new, stronger equivalence for which partial-order reductions are sound even for non-determinate processes, as well as new symmetry reductions. They demonstrate that these techniques provide a significant (several orders of magnitude) speed-up in practice, thus increasing the size of the protocols that can be analysed fully automatically. Even though the new equivalence is stronger, it is nevertheless coarse enough to avoid false attacks on most practical examples.
- The SAT-Equiv tool relies on a “small-attack property”: if there is an attack against trace equivalence, then there is a well-typed attack, that is an attack where the messages follow some a priori given structure. This allows to dramatically reduce the search space. We have recently extended [11] this approach to a class of equational theory, that encompasses all standard cryptographic primitives (including e.g. randomized encryption) as well as theories that are less considered by automatic tools, such as threshold decryption. This result will allow to further extend the SAT-Equiv tool but can also be used more generally to characterize the form of an attack, independently of the considered tool.

From a more foundational point of view, in collaboration with Erbatur (LMU, Germany) and Marshall (Univ Mary Washington, USA), Ringeissen studies decision procedures for the intruder deduction and the static equivalence problems in combinations of subterm convergent rewrite systems and syntactic theories for which it is possible to apply a mutation principle to simplify equational proofs. As a continuation of a work initially presented at UNIF'18, it has been shown that a matching property is applicable to solve both intruder deduction and static equivalence. This matching property can be satisfied when using a matching algorithm known for syntactic theories [29]. A journal paper reporting this result is currently under review.

7.1.2. Decision Procedures for Equational Theories

Participants: Christophe Ringeissen, Michaël Rusinowitch.

Equational theories and unification procedures are widely used in protocol analyzers to model the capabilities of a (passive) intruder. In the context of protocol analysis, many equational theories of practical interest satisfy the finite variant property. This class of theories is indeed a class of syntactic theories admitting a terminating mutation-based unification algorithm. This mutation-based unification algorithm generalizes the syntactic unification algorithm known for the empty theory. In collaboration with Erbatur (LMU, Germany) and Marshall (Univ Mary Washington, USA), this particular unification algorithm has been applied by Ringeissen to get new non-disjoint combination results for the unification problem [23], [32].

In collaboration with Anantharaman (LIFO, Orléans), Hibbs (SUNY Albany & Google, USA), and Narendran (SUNY Albany, USA), Rusinowitch has studied the unification problem in list theories. Decision procedures for various list theories have been investigated in the literature with applications to automated verification. In [17], it has been shown that the unifiability problem for some list theories with a *reverse* operator is NP-complete. A unifiability algorithm is given for the case where the theories are extended with a *length* operator on lists.

Among theories with the finite variant property, the class of theories presented by subterm convergent rewrite systems is particularly remarkable because it satisfies in addition a locality property. For this class of theories, it is thus possible to get a satisfiability procedure based on a reduction to the empty theory via an instantiation with the finitely many terms occurring in the input problem. As an alternative to locality, Ringeissen has investigated a politeness property, in collaboration with Chocron (Insikt Intelligence, Spain) and Fontaine (Veridis project-team). This approach has led to new non-disjoint combination results for the satisfiability problem modulo data structure theories extended with some bridging functions such as the *length* operator on lists [10], [26].

7.1.3. Recast of ProVerif

Participants: Vincent Cheval, Véronique Cortier.

Motivated by the addition of global states in ProVerif, we have started a major revision of the popular tool ProVerif. This revision goes well beyond global states and is conducted in collaboration with Bruno Blanchet, the original and main developer of ProVerif. One of the first main changes is the addition of ProVerif of the notion of "lemmas" and "axioms" that can be added to either encode additional properties (axioms) or help ProVerif to prove the desired properties. It is indeed now possible to specify lemmas, that will significantly reduce the number of considered clauses in the saturation procedure of ProVerif. These lemmas should of course be proved themselves by ProVerif, possibly by induction thanks to a particular care of the order of literals in the saturation procedure. The new approach provides more flexibility in cases where ProVerif was not able to terminate or yield false attacks (e.g. in the presence of global states).

Moreover, even when ProVerif is able to prove security, the tool is suffering from efficiency issues when applied to complex industrial protocols (up to 1 month running time for the analysis of the NoiseExplorer protocol). One reason is the subsumption procedure: a clause shall not be added if it is subsumed by another one (that is, if there exists a more general clause). This is crucial to avoid running into non termination issues. We have started a major rewrite of the subsumption procedure, taking advantage of the recent progress in this domain, in the automated deduction area. Another reason is the translation of processes into Horn clauses: For each conditional in the process, ProVerif generates a Horn clause for each possible result of this conditional.

On complex protocols with many interleaved conditionals, ProVerif is faced with an exponential blowup in the number of generated clauses. We have improved the generation of Horn clauses by avoiding exploring branches that would directly be subsumed by other conditional branches. The first experimental results show significant speed-up on many examples: On average, ProVerif is now 5 to 10 times faster than its current release, with some examples peaking at 50 to 200 times speedup.

7.1.4. Verification of Protocols with Global States

Participants: Jannik Dreier, Lucca Hirschi.

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, an extended version was accepted in the Journal of Computer Security [12].

7.1.5. Symbolic Methods in Computational Cryptography Proofs

Participants: Charlie Jacomme, Steve Kremer.

Code-based game-playing is a popular methodology for proving the security of cryptographic constructions and side-channel countermeasures. This methodology relies on treating cryptographic proofs as an instance of relational program verification (between probabilistic programs), and decomposing the latter into a series of elementary relational program verification steps. Barthe (MPI on Security and Privacy, Bochum), Grégoire (Inria SAM), Jacomme, Kremer and Strub (LIX, École Polytechnique) develop principled methods for proving such elementary steps for probabilistic programs that operate over finite fields and related algebraic structures. They focus on three essential properties: program equivalence, information flow, and uniformity. We give characterizations of these properties based on deducibility and other notions from symbolic cryptography. They use (sometimes improve) tools from symbolic cryptography to obtain decision procedures or sound proof methods for program equivalence, information flow, and uniformity. Finally, they evaluate their approach using examples drawn from provable security and from side-channel analysis - for the latter, they focus on the masking countermeasure against differential power analysis. A partial implementation of our approach is integrated in EasyCrypt, a proof assistant for provable security, and in MaskVerif, a fully automated prover for masked implementations. This work was presented at CSF [18].

7.1.6. Analysis of Deployed Protocols

Participants: Sergiu Bursuc, Lucca Hirschi, Steve Kremer.

7.1.6.1. New Privacy Threat on 3G, 4G, and Upcoming 5G AKA Protocols

Mobile communications are used by more than two-thirds of the world population who expect security and privacy guarantees. The 3rd Generation Partnership Project (3GPP) responsible for the worldwide standardization of mobile communication has designed and mandated the use of the AKA protocol to protect the subscribers' mobile services. Even though privacy was a requirement, numerous subscriber location attacks have been demonstrated against AKA, some of which have been fixed or mitigated in the enhanced AKA protocol designed for 5G.

We found and reported [9] a new privacy attack against all variants of the AKA protocol, including 5G AKA, that breaches subscriber privacy more severely than known location privacy attacks do. Our attack exploits a new logical vulnerability we uncovered that would require dedicated fixes. We demonstrate the practical feasibility of our attack using low cost and widely available setups. Finally we conduct a security analysis of the vulnerability and discuss countermeasures to remedy our attack.

Our attack has later been considered to be a *key issue in 5G [38]* by 3GPP [0]. Since then, various vendors[0] have proposed countermeasures, which are currently under discussion.

7.1.6.2. Contingent Payments

Bursuc and Kremer study protocols that rely on a public ledger infrastructure, concentrating on protocols for zero-knowledge contingent payment, whose security properties combine diverse notions of fairness and privacy. They argue that rigorous models are required for capturing the ledger semantics, the protocol-ledger interaction, the cryptographic primitives and, ultimately, the security properties one would like to achieve. Our focus is on a particular level of abstraction, where network messages are represented by a term algebra, protocol execution by state transition systems (e.g. multiset rewrite rules) and where the properties of interest can be analyzed with automated verification tools. They propose models for: (1) the rules guiding the ledger execution, taking the coin functionality of public ledgers such as Bitcoin as an example; (2) the security properties expected from ledger-based zero-knowledge contingent payment protocols; (3) two different security protocols that aim at achieving these properties relying on different ledger infrastructures; (4) reductions that allow simpler term algebras for homomorphic cryptographic schemes. Altogether, these models allow us to derive a first automated verification for ledger-based zero-knowledge contingent payment using the Tamarin prover. Furthermore, our models help in clarifying certain underlying assumptions, security and efficiency tradeoffs that should be taken into account when deploying protocols on the blockchain. This work was presented at ESORICS [20].

7.2. E-voting

7.2.1. Definitions for E-Voting

Participants: Sergiu Bursuc, Véronique Cortier, Steve Kremer, Joseph Lallemand.

Existing formal (computational) definitions for privacy in electronic voting make the assumption that the bulletin board which collects the votes behaves honestly: the only ballots on the board are created by voters, all ballots are placed without tampering with them, and no ballots are ever removed. This strong assumption is difficult to enforce in practice and whenever it does not hold vote privacy can be broken. As a consequence, voting schemes are proved secure only against an honest voting server while they are designed and claimed to resist a dishonest one. We have proposed a framework for the analysis of electronic voting schemes in the presence of malicious bulletin boards. We identify a spectrum of notions where the adversary is allowed to tamper with the bulletin board in ways that reflect practical deployment and usage considerations. To clarify the security guarantees provided by the different notions we establish a relationship with simulation-based security with respect to a family of ideal functionalities. The ideal functionalities make clear the set of authorised attacker capabilities which makes it easier to understand and compare the associated levels of security. We then leverage this relationship to show that each distinct level of ballot privacy entails some distinct form of individual verifiability. As an application, we have studied three protocols of the literature (Helios, Belenios, and Civitas) and identified the different levels of privacy they offer. This work has appeared as a part of the PhD thesis [8], defended by Joseph Lallemand in November 2019.

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 [40]. This result has been presented at EuroS&P [19].

7.2.2. Design of E-Voting Protocols

Participants: Véronique Cortier, Jannik Dreier, Joseph Lallemand, Mathieu Turuani.

[0] 3rd Generation Partnership Project, responsible for the standardization of 3G, 4G, and 5G mobile networks

[0] Qualcomm, Gemalto, China Mobile, Mobile Thales Thales, Nokia Nokia, ZTE ZTE, and Huawei.

Most existing voting systems either assume trust in the voting device or in the voting server. Filipiak (Orange Labs), 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 [42]. This contribution is of independent interest. This work has been presented at CSF'19 [22].

As a part of a contract with Idemia, we are designing a novel electronic voting system tailored to their needs. The system is made for on-site elections, with the use of smart cards. However, the goal is that the trust should not be placed in one single part of the system, hence smart cards can not be trusted. One originality of the approach is the possibility to re-use existing techniques, in conjunction with the use of smart-cards and paper ballots. In this context, we have designed a novel audit technique [36], which can be seen as a variant to the "cast or audit" approach proposed by Josh Benaloh. One significant advantage of our solution is that voters now audit systematically their ballot (instead of choosing whether they should audit or not) and cast the audited ballot.

7.3. Online Social Networks

7.3.1. Privacy Protection in Social Networks

Participants: Bizhan Alipour, Abdessamad Imine, Michaël Rusinowitch.

Social media such as Facebook provides a new way to connect, interact and learn. Facebook allows users to share photos and express their feelings by using comments. However, Facebook users are vulnerable to attribute inference attacks where an attacker intends to guess private attributes (e.g., gender, age, political view) of target users through their online profiles and/or their vicinity (e.g., what their friends reveal). Given user-generated pictures on Facebook, we show in [16] how to launch gender inference attacks on their owners from pictures meta-data composed of: (i) alt-texts generated by Facebook to describe the content of pictures, and (ii) comments posted by friends, friends of friends or regular users. We assume these two meta-data are the only available information to the attacker. Evaluation results demonstrate that our attack technique can infer the gender with an accuracy of 84% by leveraging only alt-texts, 96% by using only comments, and 98% by combining alt-texts and comments. We compute a set of sensitive words that enable attackers to perform effective gender inference attacks. We show the adversary prediction accuracy is decreased by hiding these sensitive words. To the best of our knowledge, this is the first inference attack on Facebook that exploits comments and alt-texts solely. In subsequent work we have investigated the case where comments are reduced to Emojis.

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 Numeryx company, 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 [33], [34].

PRIVATICS Project-Team

6. New Results

6.1. Differential Inference Testing

Participant: Claude Castelluccia.

In order to protect individuals' privacy, data have to be "well-sanitized" before sharing them, i.e. one has to remove any personal information before sharing data. However, it is not always clear when data shall be deemed well-sanitized. In [10], we argue that the evaluation of sanitized data should be based on whether the data allows the inference of sensitive information that is specific to an individual, instead of being centered around the concept of re-identification. We propose a framework to evaluate the effectiveness of different sanitization techniques on a given dataset by measuring how much an individual's record from the sanitized dataset influences the inference of his/her own sensitive attribute. Our intent is not to accurately predict any sensitive attribute but rather to measure the impact of a single record on the inference of sensitive information. We demonstrate our approach by sanitizing two real datasets in different privacy models and evaluate/compare each sanitized dataset in our framework.

6.2. Analyse des impacts de la reconnaissance faciale - Quelques éléments de méthode (in French)

Participants: Claude Castelluccia, Daniel Le Métayer.

Significant technical progress has been made in recent years in the field of image processing, in particular in facial recognition. The deployments and experiments of this type of systems are more and more numerous. However, opinions differ on their use, especially in public space. Noting the lack of consensus on a technology that can have a significant impact on society, many organizations have alerted public opinion and asked for a public debate on the subject. We believe that such a debate is indeed necessary. However, for it to be truly productive, it is necessary to be able to confront the arguments in a rigorous manner while avoiding, as far as possible, the preconceptions, and by distinguishing established facts from assumptions or opinions. The purpose of this document [14] is precisely to help put the terms of the debate on solid foundations. It is therefore not a question here of taking a position on facial recognition in general nor of providing an exhaustive review of its applications but of proposing elements of method, illustrated by a few examples. We first present a quick overview of the applications of facial recognition before detailing the reasons that make it a particularly sensitive subject, emphasizing in particular the risks linked to a possible generalization of its use. We then present an incremental, comparative and rigorous approach to analyze the impacts of a facial recognition system.

6.3. Towards a generic framework for black-box explanation methods

Participants: Daniel Le Métayer, Clément Hénin.

Explainability has generated increased interest during the last decade because the most accurate ML techniques often lead to opaque Algorithmic Decision Systems (ADS) and opacity is a major source of mistrust. Indeed, even if explanations are not a panacea, they can play a key role, not only to enhance trust in the system, but also to allow its users to better understand its outputs and therefore to make a better use of it. In addition, they are necessary to make it possible to challenge the decisions resulting from an ADS. Explanations can take different forms, they can target different types of users and different types of methods can be used to produce them. Our work on this topic [15] focuses on a category of methods, called "black-box", that do not make any assumption about the availability of the code of the ADS or its implementation techniques. Our first contribution is to bring to light a common structure for Black-box Explanation Methods and to define a generic framework allowing us to compare and classify different approaches. This framework consists of three components, called respectively

Sampling, Generation and Interaction. Beyond its interest as a systematic presentation of the state of the art, we believe that this framework can also provide new insights for the design of new explanation systems. For example, it may suggest new combinations of Sampling and Generation components or criteria to choose the most appropriate combination to produce a given type of explanation.

6.4. A generic information and consent framework for the IoT

Participants: Daniel Le Métayer, Mathieu Cunche, Victor Morel.

The development of the Internet of Things (IoT) raises specific privacy issues especially with respect to information and consent. People are generally unaware of the devices collecting data about them and do not know the organizations operating them. Solutions such as stickers or wall signs are not effective information means in most situations. As far as consent is concerned, individuals do not have simple means to express and communicate it to the entities collecting data. Furthermore, the devices used to collect data in IoT environments have scarce resources; some of them do not have any user interface, are battery-operated or operate passively. The Working Party 29 (now "European Data Protection Board") advocates the design of new consent mechanisms, such as "privacy proxies", on the devices themselves. Starting from their recommendations, we have defined general requirements that have to be met to ensure that information and consent are managed in a manner that is satisfactory both for data subjects and for data controllers. We have shown in [8] how these requirements can be implemented in different situations, in particular through declaration registers and beacons. Depending on the context and the types of devices involved, not all technical options are always possible. In order to provide guidance to IoT system designers, we have outlined the main choice factors in the design pace are illustrated the framework with several challenging case studies. We have also implemented a Proof of Concept prototype implementation of these techniques.

6.5. Analysis of privacy policies to enhance informed consent

Participant: Daniel Le Métayer.

A privacy policy language must meet a number of requirements to be able to express the valid consent of the data subject for the processing of their personal data. For example, under the GDPR, valid consent must be freely given, specific, informed and unambiguous. Therefore, the language must be endowed with a formal semantics in order to avoid any ambiguity about the meaning of a privacy policy. However, the mere existence of a semantics does not imply that DSs properly understand the meaning of a policy and its potential consequences. One way to enhance the understanding of the data subjects is to provide them information about the potential risks related to a privacy policy. This is in line with Recital 39 of the GDPR which stipulates that data subjects should be "made aware of the risks, rules, safeguards and rights in relation to the processing of personal data and how to exercise their rights in relation to such processing". To address this need, we have defined a language in [11], called PILOT, meeting these requirements and shown its benefits to define precise privacy policies and to highlight the associated privacy risks. In order to automatically answer questions related to privacy risks, we use the verification tool SPIN and the modeling language PROMELA. Risk properties are encoded in Linear Temporal Logic properties that can be automatically checked by SPIN.

6.6. Understanding algorithmic decision-making: Opportunities and challenges, Study for the European Parliament (STOA)

Participants: Claude Castelluccia, Daniel Le Métayer.

Algorithms are far from being a recent invention but they are increasingly involved in systems used to support decision making. Algorithmic Decision Systems (ADS) often rely on the analysis of large amounts of personal data to infer correlations or, more generally, to derive information deemed useful to make decisions. Humans may have a role of varying degree in the decision making and may even be completely out of the loop in entirely automated systems. In many situations, the impact of the decision on people can be significant: access to credit, employment, medical treatment, judicial sentences, etc. Entrusting ADS to make or to influence such decisions raises a variety of issues that differ in nature such as ethical, political, legal, technical, etc. and great care must be taken to analyse and address these issues. If they are neglected, the expected benefits of these systems may be offset by the variety of risks for individuals (discrimination, unfair practices, loss of autonomy, etc.), the economy (unfair practices, limited access to markets, etc.) and society as a whole (manipulation, threat to democracy, etc.).

We have written a report for the European Parliament reviewing the opportunities and risks related to the use of ADS. We present existing options to reduce these risks and explain their limitations. We sketch some recommendations to benefit from the tremendous possibilities of ADS while limiting the risks related to their use. Beyond providing an up-to-date and systematic review of the situation, the report gives a precise definition of a number of key terms and an analysis of their differences. This helps clarify the debate. The main focus of the report is the technical aspects of ADS. However, other legal, ethical and social dimensions are considered to broaden the discussion.

6.7. Saving Private Addresses: An Analysis of Privacy Issues in the Bluetooth-Low-Energy Advertising Mechanism

Participants: Mathieu Cunche, Guillaume Celiosa.

The Bluetooth Low Energy (BLE) protocol is being included in a growing number of connected objects such as fitness trackers and headphones. As part of the service discovery mechanism of BLE, devices announce themselves by broadcasting radio signals called advertisement packets that can be collected with off-the-shelf hardware and software. To avoid the risk of tracking based on those messages, BLE features an address randomization mechanism that substitutes the device address with random temporary pseudonyms,called Private addresses. We analyze the privacy issues associated with the advertising mechanism of BLE, leveraging a large dataset of advertisement packets collected in the wild. First, we identified in [7] that some implementations fail at following the BLE specifications on the maximum lifetime and the uniform distribution of random identifiers. Furthermore, we found that the payload of the advertisement packet can hamper the randomization mechanism by exposing counters and static identifiers. In particular, we discovered that advertising data of Apple and Microsoft proximity protocols can be used to defeat the address randomization scheme. Finally, we discuss how some elements of advertising data can be leveraged to identify the type of device, exposing the owner to inventory attacks

6.8. Fingerprinting Bluetooth-Low-Energy Devices Based on the Generic Attribute Profile

Participants: Mathieu Cunche, Guillaume Celiosa.

Bluetooth Low Energy (BLE) is a short range wireless technology included in many consumer devices such as smartphones, earphones and wristbands. As part of the Attribute (ATT) protocol, discoverable BLE devices expose a data structure called Generic Attribute (GATT) profile that describes supported features using concepts of services and characteristics. This profile can be accessed by any device in range and can expose users to privacy issues. We study how the GATT profile can be used to create a fingerprint that can be exploited to circumvent anti-tracking features of the BLE standard (i.e. MAC address randomization). Leveraging a dataset of more than 13000 profiles, we analyze the potential of this fingerprint and show that it can be used to uniquely identify a number of devices. We also shed light in [6] on several issues where GATT profiles can be mined to infer sensitive information that can impact privacy of users. Finally, we suggest solutions to mitigate those issues.

6.9. Privacy implications of switching ON a light bulb in the IoT world

Participants: Vincent Roca, Mathieu Thiery.

The number of connected devices is increasing every day, creating smart homes and shaping the era of the Internet of Things (IoT), and most of the time, end-users are unaware of their impacts on privacy. We analyze in [23] the ecosystem around a Philips Hue smart white bulb in order to assess the privacy risks associated to the use of different devices (smart speaker or button) and smartphone applications to control it. We show that using different techniques to switch ON or OFF this bulb has significant consequences regarding the actors involved (who mechanically gather information on the user's home) and the volume of data sent to the Internet (we measured differences up to a factor 100, depending on the control technique we used). Even when the user is at home, these data flows often leave the user's country, creating a situation that is neither privacy friendly (and the user is most of the time ignorant of the situation), nor sovereign (the user depends on foreign actors), nor sustainable (the extra energetic consumption is far from negligible). We therefore advocate a complete change of approach, that favors local communications whenever sufficient.

6.10. Security Analysis of Subject Access Request Procedures How to authenticate data subjects safely when they request for their data

Participants: Cédric Lauradoux, Coline Boniface.

With the GDPR in force in the EU since May 2018, companies and administrations need to be vigilant about the personal data they process. The new regulation defines rights for data subjects and obligations for data controllers but it is unclear how subjects and controllers interact concretely. In [4], we try to answer two critical questions: is it safe for a data subject to exercise the right of access of her own data? When does a data controller have enough information to authenticate a data subject? To answer these questions, we have analyzed recommendations of Data Protection Authorities and authentication practices implemented in popular websites and third-party tracking services. We observed that some data controllers use unsafe or doubtful procedures to authenticate data subjects. The most common flaw is the use of authentication based on a copy of the subject's national identity card transmitted over an insecure channel. We define how a data controller should react to a subject's request to determine the appropriate procedures to identify the subject and her data. We provide compliance guidelines on data access response procedures.

6.11. Plausible Deniability for Practical Privacy-Preserving Live Streaming

Participant: Antoine Boutet.

Video consumption is one of the most popular Internet activities worldwide. The emergence of sharing videos directly recorded with smartphones raises important privacy concerns. In this work we propose P3LS, the first practical privacy-preserving peer-to-peer live streaming system. To protect the privacy of its users, P3LS relies on k-anonymity when users subscribe to streams, and on plausible deniability for the dissemination of video streams. Specifically, plausible deniability during the dissemination phase ensures that an adversary is never able to distinguish a user's stream of interest from the fake streams from a statistical analysis (i.e., using an analysis of variance). We exhaustively evaluate P3LS and show that adversaries are not able to identify the real stream of a user with very high confidence. Moreover, P3LS consumes 30% less bandwidth than the standard k-anonymity approach where nodes fully contribute to the dissemination of k streams.

6.12. Protecting motion sensor data against sensitive inferences through an adversarial network approach

Participants: Antoine Boutet, Théo Jourdan.

With the widespread development of the quantified self movement, more and more motion sensor data are captured and transmitted through the intermediary of smartphones. However, granting to applications a direct access to sensor data expose users to many privacy risks, including in particular the possibility of inferring their activities and transportation mode to more sensitive inferences such as their demographic attributes or even mobility deficiency. In this work, we propose a privacy-preserving scheme to protect sensor data for activity recognition while at the same time preventing unwanted sensitive inferences on specific information. To achieve this objective, we leverage on the powerful framework of generative adversarial networks (GANs) to sanitize the sensor data. More precisely in our framework three neural networks are jointly trained, a generator that aim at sanitizing the data given at input as well two discriminators that try to infer respectively the sensitive attributes and the current activity of the user. By letting these neural networks compete against each other, the mechanism improves the protection while providing a good accuracy in terms of activity recognition and limiting sensitive inferences on specified attributes. Preliminary results demonstrate that the approach is promising in terms of achieving a good utility-privacy trade-off.

6.13. Inria white book on Cybersecurity: Current challenges and Inria's research directions

Participant: Vincent Roca.

This book provides an overview of research areas in cybersecurity, illustrated by contributions from Inria teams. The first step in cybersecurity is to identify threats and define a corresponding attacker model. Threats, including malware, physical damage or social engineering, can target the hardware, the network, the operating system, the applications, or the users themselves.

Then, detection and protection mechanisms must be designed to defend against these threats. One of the core mechanisms is cryptography, in order to ensure the confidentiality and integrity of data. These primitives must be the object of continuous cryptanalysis to ensure the highest level of security. However, secure cryptographic primitives alone are not sufficient for secure communications and services: cryptographic protocols, implementing richer interactions on top of the primitives, are needed. These protocols are distributed systems. Ensuring that they achieve their goals in the presence of an adversary requires the use of formal verification techniques, which have been extremely successful in this field.

Additional security services, such as authentication and access control, are needed to enforce a security policy. These security services, usually provided by the operating system or the network devices, can themselves be attacked and sometimes bypassed. Therefore, activities on the information system are monitored in order to detect any violation of the security policy. Finally, as attacks can spread extremely fast, the system must react automatically or at least reconfigure itself to avoid propagating attacks.

Privacy has also become an intrinsic part of cybersecurity. Privacy has its own properties, techniques, and methodology. Moreover, the study of privacy often requires to take legal, economical, and sociological aspects into account.

All these security mechanisms need to be carefully integrated in security-critical applications. These applications include traditional safety-critical applications that are becoming increasingly connected and therefore more vulnerable to security attacks, as well as new infrastructures running in the cloud or connected to a multitude of Things (IoT).

6.14. Inspect what your location history reveals about you - Raising user awareness on privacy threats associated with disclosing his location data

Participant: Antoine Boutet.

Location is one of the most extensively collected personal data on mobile by applications and third-party services. However, how the location of users is actually processed in practice by the actors of targeted advertising ecosystem remains unclear. Nonetheless, these providers have a strong incentive to create very detailed profile of users to better monetize the collected data. End users are usually not aware about the strength and wide range of inference that can be performed from their mobility traces. In this work, users interact with a web-based application to inspect their location history and to discover the inferential power of this kind of data. Moreover to better understand the possible countermeasures, users can apply a sanitization to protect their data and visualize the impact on both the mobility traces and the associated inferred information. The objective of this work is to raise the user awareness on the profiling capabilities and the privacy threats associated with disclosing his location data as well as how sanitization mechanisms can be efficient to mitigate these privacy risks. In addition, by collecting users feedbacks on the personal information revealed and the usage of a geosanitization mechanism, we hope that this work will also be useful to constitute a new and valuable dataset on users perceptions on these questions.

6.15. Pseudonymisation techniques and best practices

Participant: Cédric Lauradoux.

This ENISA report explores further the basic notions of pseudonymisation, as well as technical solutions that can support implementation in practice. Starting from a number of pseudonymisation scenarios, the report defines first the main actors that can be involved in the process of pseudonymisation along with their possible roles. It then analyses the different adversarial models and attacking techniques against pseudonymisation, such as brute force attack, dictionary search and guesswork. Moreover, it presents the main pseudonymisation techniques and policies available today.

PROSECCO Project-Team

7. New Results

7.1. Verification of security protocols

Participants: Bruno Blanchet, Karthikeyan Bhargavan, Benjamin Lipp.

Our verification of the WireGuard open source Virtual Private Network (VPN) with CryptoVerif appears at EuroS&P 2019 [22], [27].

We continued the development of our protocol verification tools ProVerif and CryptoVerif. The new features of this year are detailed in the section on software.

In the setting of the ANR AnaStaSec project, we worked on the verification of avionic security protocols. More specifically, in 2015, we had verified the protocol of the Secure Dialog Service using ProVerif and CryptoVerif and recommended many changes to the specification. The ICAO started to take into account our remarks, and this year we analyzed a new version of the specification. Our analysis showed that many recommendations still need to be taken into account. Additionally, we also commented on the recent choice of using DTLS over UDP to secure the future ATN/IPS (Aeronautical Telecommunication Network / Internet Protocol Suite) network, which seems very positive. The details of these results are still confidential; they have been provided to ANR.

7.2. Verified Software for Cryptographic Web Applications

Participants: Karthikeyan Bhargavan, Benjamin Beurdouche, Denis Merigoux, Jonathan Protzenko.

WebAssembly in a new language runtime that is now supported by all major web browsers and web application frameworks. We developed a compiler from the Low* subset of the F* programming language to WebAssembly and used this compiler to translate our HACL* verified cryptographic library to WebAssembly, hence obtaining the first verified cryptographic library for the Web. We also used this framework to develop and verify an implementation of the Signal protocol in WebAssembly, and demonstrated how this implementation can be used as a drop-in replacement for the libsignal-protocol library used in mainstream messaging applications like Signal, WhatsApp, and Skype.

Our work was published at the IEEE Security and Privacy conference [24]. Our WebAssembly version of HACL* and our verified Signal implementation were publicly released as open source on GitHub.

7.3. Journey beyond full abstraction

Participants: Carmine Abate, Roberto Blanco, Deepak Garg [MPI-SWS], Catalin Hritcu, Marco Patrignani [Stanford and CISPA], Jérémy Thibault.

Even for safe languages, all guarantees are lost when interacting with low-level code, for instance when using low-level libraries. A compromised or malicious library that gets linked in can easily read and write data and code, jump to arbitrary instructions, or smash the stack, blatantly violating any source-level abstraction and breaking any guarantee obtained by source-level reasoning. Our goal is to build formally secure compartmentalizing compilation chains that defend against such attacks. We started by investigating what it means for a compilation chain to provide secure interoperability between a safe source language and linked target-level code that is adversarial. In this model, a secure compilation chain ensures that even linked adversarial target-level code cannot break the security properties of a compiled program any more than some linked source-level code could. However, the precise class of security properties one chooses to preserve crucially impacts not only the supported security goals and the strength of the attacker model, but also the kind of protections the compilation chain has to introduce and the kind of proof techniques one can use to make sure that the protections are watertight. We are the first to thoroughly explore a large space of secure compilation criteria based on the preservation against adversarial contexts of various classes of trace properties such as safety, of hyperproperties such as noninterference, and of relational hyperproperties such as trace equivalence [17], [10].

7.4. Principles of Program Verification for Arbitrary Monadic Effects

Participants: Kenji Maillard, Danel Ahman [University of Ljubljana], Robert Atkey [University of Strathclyde], Guido Martinez, Catalin Hritcu, Exequiel Rivas, Éric Tanter, Antoine Van Muylder, Cezar Andrici.

We devised a principled semantic framework for verifying programs with arbitrary monadic effects in a generic way with respect to expressive specifications. The starting point are Dijkstra monads, which are monad-like structures that classify effectful computations satisfying a specification drawn from a monad. Dijkstra monads have already proven valuable in practice for verifying effectful code, and in particular, they allow the F* program verifier to compute verification conditions.

We provide the first semantic investigation of the algebraic structure underlying Dijkstra monads [13], [11] and unveil a close relationship between Dijkstra monads and effect observations, i.e., mappings between a computational and a specification monad that respect their monadic structure. Effect observations are flexible enough to provide various interpretations of effects, for instance total vs partial correctness, or angelic vs demonic nondeterminism. Our semantic investigation relies on a general theory of specification monads and effect observations, using an enriched notion of relative monads and relative monad morphisms. We moreover show that a large variety of specification monads can be obtained by applying monad transformers to various base specification monads, including predicate transformers and Hoare-style pre- and postconditions. For defining correct monad transformers, we design a language inspired by the categorical analysis of the relationship between monad transformers and algebras for a monad.

We also adapt our framework to relational verification [14], [11], i.e., proving relational properties between multiple runs of one or more programs, such as noninterference or program equivalence. For this we extend specification monads and effect observations to the relational setting and use them to derive the semantics and core rules of a relational program logic generically for any monadic effect. Finally, we identify and overcome conceptual challenges that prevented previous relational program logics from properly dealing with effects such as exceptions, and are the first to provide a proper semantic foundation and a relational program logic for exceptions.

7.5. 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*[23], 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.

TAMIS Project-Team

6. New Results

6.1. Results for Axis 1: Vulnerability analysis

6.1.1. New Advances on Side-channel Distinguishers

Participants: Christophe Genevey Metat, Annelie Heuser.

A Systematic Evaluation of Profiling Through Focused Feature Selection.
Profiled side-channel attacks consist of several steps one needs to take. An important, but sometimes ignored, step is a selection of the points of interest (features) within side-channel measurement traces. A large majority of the related works start the analyses with an assumption that the features are preselected. Contrary to this assumption, here, we concentrate on the feature selection step. We investigate how advanced feature selection techniques stemming from the machine learning domain can be used to improve the attack efficiency. To this end, we provide a systematic evaluation of the methods of interest. The experiments are performed on several real-world data sets containing software and hardware implementations of AES, including the random delay countermeasure. Our results show that wrapper and hybrid feature selection methods perform extremely well over a wide range of test scenarios and a number of features selected. We emphasize L1 regularization (wrapper approach) and linear support vector machine (SVM) with recursive feature elimination used after chi-square filter (Hybrid approach) that performs well in both accuracy and guessing entropy. Finally, we show that the use of appropriate feature selection techniques is more important for an attack on the high-noise data sets, including those with countermeasures, than on the low-noise ones.

[3] Make Some Noise. Unleashing the Power of Convolutional Neural Networks for Profiled Side-channel Analysis. *Profiled side-channel analysis 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 able to reach high performance for a number of considered datasets. We compare our neural network with the one designed for a particular dataset with masking countermeasure and we show that both are good designs but also that 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 reduce the number of measurements needed to reveal the secret key by orders of magnitude for both neural networks. Our new convolutional neural network instance with added noise is able to break the implementation protected with the random delay countermeasure by using only 3 traces in the attack phase. To further strengthen our experimental results, we investigate the performance with a varying number of training samples, noise levels, and epochs. Our findings show that adding noise is beneficial throughout all training set sizes and epochs.*

The Curse of Class Imbalance and Conflicting Metrics with Machine Learning for Side-channel Evaluations.

We concentrate on machine learning techniques used for profiled sidechannel 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.

[5] CC Meets FIPS: A Hybrid Test Methodology for First Order Side Channel Analysis.
Common Criteria (CC) and FIPS 140-3 are two popular side channel testing methodologies. Test Vector Leakage Assessment Methodology (TVLA), a potential candidate for FIPS, can detect the presence of side-channel information in leakage measurements. However, TVLA results cannot be used to quantify side-channel vulnerability and it is an open problem to derive its relationship with side channel attack success rate (SR), i.e., a common metric for CC. In this paper, we extend the TVLA testing beyond its current scope. Precisely, we derive a concrete relationship between TVLA and signal to noise ratio (SNR). The linking of the two metrics allows direct computation of success rate (SR) from TVLA for given choice of intermediate variable and leakage model and thus unify these popular side channel detection and evaluation metrics. An end-to-end methodology is proposed, which can be easily automated, to derive attack SR starting from TVLA testing. The methodology works under both univariate and multivariate setting and is capable of quantifying any first order leakage. Detailed experiments have been provided using both simulated traces and real traces on SAKURA-GW platform. Additionally, the proposed methodology is benchmarked against previously published attacks on DPA contest v4.0 traces, followed by extension to jitter based countermeasure. The result shows that the proposed methodology provides a quick estimate of SR without performing actual attacks, thus bridging the gap between CC and FIPS.

[13] Combining sources of side-channel information.
A few papers relate that multi-channel consideration can be beneficial for side-channel analysis. However, all were conducted using classical attack techniques. In this work, we propose to explore a multi-channel approach thanks to machine/deep learning. We investigate two kinds of multi-channel combinations. Unlike previous works, we investigate the combination of EM emissions from different locations capturing data-dependent leakage information on the device. Additionally, we consider the combination of the classical leaking signals and a measure of mostly the ambient noise. The knowledge of the ambient noise (due to WiFi, GSM, ...) may help to remove it from a noisy trace. To investigate these multi-channel approaches, we describe one option of how to extend a CNN architecture which takes as input multiple channels. Our results show that multi-channel networks are suitable for side-channel analysis. However, if one channel alone already contains enough exploitable information to reach high effectiveness, naturally, the multi-channel approach cannot improve the performance further.

6.1.2. Side-channel analysis on post-quantum cryptography

Participants: Tania Richmond, Yulliwas Ameur, Agathe Cheriere, Annelie Heuser.

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, in particular code-based cryptosystems.

During this year, we have set up a first side-channel experiment platform suited for embedded devices running code-based cryptosystems. Using this platform we exploited vulnerabilities of the syndrome computation present in some code-based algorithms.

6.1.3. Verification of IKEv2 protocol

Participants: Tristan Ninet, Olivier Zendra.

The IKEv2 (Internet Key Exchange version 2) protocol is the authenticated key-exchange protocol used to set up secure communications in an IPsec (Internet Protocol security) architecture. 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 explored more in-depth a formal technique and work towards the goal of verifying generic properties (absence of implementation flaws) on softwares like strongSwan.

Publications:

- [10] Model Checking the IKEv2 Protocol Using Spin
- [11] The Deviation Attack: A Novel Denial-of-Service Attack Against IKEv2

6.1.4. Software obfuscation

Participants: Alexandre Gonzalvez, Olivier Decourbe.

The limits of software obfuscation are not clear in practice. A protection based on opaque predicates can not be compatible with the control flow integrity property at low-level, due to the presence of indirect jumps in the instruction set architecture semantics. We propose a restricted instruction set architecture to overcome this limit. We argue for the adoption of restricted instruction set architecture for security-related computation.
Publication:

- [9] A case against indirect jumps for secure programs

6.2. 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:

6.2.1. Malware Classification and clustering

Participants: Cassius Puodzius, Stefano Sebastio, Olivier Decourbe, Annelie Heuser, Olivier Zendra.

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.

During this year we build a end-to-end functional toolchain for supervised learning.

Furthermore, we have extended our approach to malware classification using unsupervised clustering. Preliminary results show that we are able to classify malware according to their behavioral properties without the need of any predefined labels.

6.2.2. *Packers analysis*

Participants: Lamine Nourredine, Cassius Puodzius, Stefano Sebastio, Annelie Heuser, Olivier Zendra.

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.

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.

During this year we have developed a toolchain of effective clustering of packers, in particular taking into account the possibility of evolution in packers. For this we derived and implemented new feature extraction strategies combined with incremental clustering algorithms.

6.3. (Coordination of the) H2020 TeamPlay Project, and Expression of Security Properties

Participants: Olivier Zendra, Yoann Marquer, Céline Minh, Nicolas Kiss, Annelie Heuser, Tania Richmond.

6.3.1. Overview & results

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 “ladderized" version with higher security level.

6.3.2. Publication

- Type-Driven Verification of Non-functional Properties [8].