IN PARTNERSHIP WITH:
Université Rennes 1

Activity Report 2018

Project-Team WIDE

The World Is DistributEd: Exploring the tension between scale and coordination

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed Systems and middleware

Table of contents

1. Team, Visitors, External Collaborators **1**
2. Overall Objectives **2**
2.1. Overview 2
2.2. Planetary-Scale Geo-Distributed Systems 2
2.3. Highly Personalized On-Line Services 3
2.4. Social Collaboration Platforms 3
3. Research Program **4**
3.1. Overview 4
3.2. Hybrid Scalable Architectures 5
3.2.1. Browser-based fog computing 5
3.2.2. Emergent micro-service deployment and management 6
3.3. Personalizable Privacy-Aware Distributed Systems 7
3.3.1. Privacy-preserving decentralized learning 7
3.3.2. Personalization in user-oriented applications 7
3.3.3. Privacy preserving decentralized aggregation 7
3.4. Network Diffusion Processes 8
3.4.1. Spread maximization 8
3.4.2. Immunization optimization 9
3.5. Systemizing Modular Distributed Computability and Efficiency 9
3.5.1. Randomized algorithm for mutual exclusion and coordination 9
3.5.2. Modular theory of distributed computing 10
4. Highlights of the Year **10**
5. New Software and Platforms **10**
5.1. WebGC 10
5.2. YALPS 11
5.3. KIFF 11
6. New Results **11**
6.1. Scalable Systems 11
6.1.1. Nobody cares if you liked Star Wars: KNN graph construction on the cheap. 11
6.1.2. Pleiades: Distributed structural invariants at scale 12
6.1.3. CASCADE: Reliable distributed session handoff for continuous interaction across devices 12
6.1.4. Sprinkler: A probabilistic dissemination protocol to provide fluid user interaction in multi-device ecosystems 12
6.2. Personalization and Privacy 13
6.2.1. GoldFinger 13
6.2.2. Collaborative filtering under a Sybil attack: Similarity metrics do matter! 13
6.3. Network and Graph Algorithms 13
6.3.1. Rumor spreading and conductance 13
6.3.2. Tight bounds for coalescing-branching random walks on regular graphs 13
6.3.3. The quadratic shortest path problem: Complexity, approximability, and solution methods 14
6.3.4. Weighting past on the geo-aware state deployment problem 14
6.3.5. Mind the gap: Autonomous detection of partitioned MANET systems using opportunistic aggregation 14
6.4. Theory of Distributed Systems 15
6.4.1. An improved bound for random binary search trees with concurrent insertions 15
6.4.2. Acyclic strategy for silent self-stabilization in spanning forests 15
6.4.3. Set agreement and renaming in the presence of contention-related crash failures 16

6.4.4. Anonymous obstruction-free (n, k)-set agreement with $n-k+1$ atomic read/write registers 16
7. Bilateral Contracts and Grants with Industry **17**
7.1. Bilateral Contracts with Industry 17
7.2. Bilateral Grants with Industry 17
8. Partnerships and Cooperations **17**
8.1. Regional Initiatives 17
8.2. National Initiatives 17
8.2.1. Labex CominLab Descent (2013-2018) 17
8.2.2. ANR Project PAMELA (2016-2020) 18
8.2.3. ANR Project OBrowser (2016-2020) 18
8.2.4. ANR Project DESCARTES (2016-2020) 18
8.2.5. ANR-ERC Tremplin Project NDFUSION (2016-2018) 18
8.2.6. Labex CominLab PROFILE (2016-2019) 18
8.3. International Initiatives 19
8.4. International Research Visitors 19
8.4.1. Visits of International Scientists 19
8.4.2. Visits to International Teams 19
9. Dissemination **19**
9.1. Promoting Scientific Activities 19
9.1.1. Scientific Events Organisation 19
9.1.2. Scientific Events Selection 20
9.1.2.1. Chair of Conference Program Committees 20
9.1.2.2. Member of the Conference Program Committees 20
9.1.3. Journal 20
9.1.4. Invited Talks 20
9.1.5. Leadership within the Scientific Community 21
9.1.6. Research Administration 21
9.2. Teaching - Supervision - Juries 21
9.2.1. Teaching 21
9.2.2. Supervision 22
9.2.3. Juries 22
9.3. Popularization 22
10. Bibliography **23**

Project-Team WIDE

Creation of the Team: 2018 January 01, updated into Project-Team: 2018 June 01

Keywords:

Computer Science and Digital Science:

A1.2.5. - Internet of things
A1.2.9. - Social Networks
A1.3.2. - Mobile distributed systems
A1.3.3. - Blockchain
A1.3.4. - Peer to peer
A1.3.5. - Cloud
A1.3.6. - Fog, Edge
A2.1.7. - Distributed programming
A2.6.2. - Middleware
A2.6.3. - Virtual machines
A3.5.1. - Analysis of large graphs
A4.8. - Privacy-enhancing technologies
A7.1.1. - Distributed algorithms
A7.1.2. - Parallel algorithms
A7.1.3. - Graph algorithms
A9.2. - Machine learning

Other Research Topics and Application Domains:

B6.1.1. - Software engineering
B6.3.1. - Web
B6.3.5. - Search engines
B6.4. - Internet of things
B9.5.1. - Computer science
B9.5.6. - Data science

1. Team, Visitors, External Collaborators

Research Scientists

Davide Frey [Inria, Researcher]
George Giakkoupis [Inria, Researcher]
Erwan Le Merrer [Inria, Advanced Research Position, from Nov 2018, HDR]

Faculty Members

David Bromberg [Univ de Rennes I, Professor, HDR]
Michel Raynal [Univ de Rennes I, Emeritus, HDR]
François Taïani [Team leader, Univ de Rennes I, Professor, HDR]

Post-Doctoral Fellow

Anaïs Durand [Univ de Rennes I, from Apr 2018 until Oct 2018]

PhD Students

Loïck Bonniot [Technicolor, from May 2018]

Amaury Bouchra Pilet [Univ de Rennes I, from Sep 2018]
Simon Bouget [Univ de Rennes I, until Sep 2018]
Quentin Dufour [Inria]
Louison Gitzinger [Univ de Rennes I]
Adrien Luxey [Univ de Rennes I]
Pierre-Louis Roman [Univ de Rennes I]
Olivier Ruas [Inria]

Interns

Alex Auvolat Bernstein [Ecole Normale Supérieure Paris, from Sep 2018]
Dylan Bellier [Univ de Rennes I, from May 2018 until Jul 2018]
Amaury Bouchra Pilet [Inria, from Mar 2018 until Jul 2018]
Nominoe Kervadec [Univ de Rennes I, from Jul 2018 until Aug 2018]
Mamane Mati Ibrahim [Univ de Rennes I, from Jun 2018 until Sep 2018]

Administrative Assistant

Virginie Desroches [Inria]

Visiting Scientists

Arvid Bosk [KTH Royal Institute of Technology, until Jun 2018]
Fabio Costa [Federal University of Goiás, from Sep 2018]
Paulo Ferreira [INESC ID Lisboa, from Feb 2018 until Mar 2018]

2. Overall Objectives

2.1. Overview

The long term goal of the WIDE team is to provide the practical tools and theoretical foundations required to address the scale, dynamicity, and uncertainty that constitute the foundations of modern distributed computer systems. In particular, we would like to **explore the inherent tension between scalability and coordination guarantees**, and develop novel techniques and paradigms that are adapted to the rapid and profound changes impacting today's distributed systems, both in terms of the application domains they support and the operational constraints they must meet.

These changes are particularly visible in three key areas related to our research: *(i)* planetary-scale information systems, *(ii)* personalized services, and *(iii)* new forms of social applications (e.g. in the field of the sharing economy).

2.2. Planetary-Scale Geo-Distributed Systems

Modern large-scale systems often encompass thousands of server nodes, hosted in tens of datacenters distributed over several continents. To address the challenges posed by such systems, alternative distributed architectures are today emerging that emphasize *decentralized* and *loosely coupled* interactions. This evolution can be observed at multiple levels of an application's distributed stack: the growing interest, both practical and theoretical, for weak consistency models is such an example. In spite of their potential counter-intuitive behaviors, weakly consistent data-structures allow developers to trade strict coordination guarantees for the ability to deliver a reactive and scalable service even when hit by arbitrary network delays or system partitions. At a higher, more architectural level, similar motivations explain the push for *micro-services* on the server side of on-line applications and the growth of rich *browser-based programming technologies* on their client side. Micro services help development teams decompose complex applications into a set of simpler and loosely-connected distributed services. In a parallel evolution, modern browsers embark increasingly powerful networking APIs such as WebRTC. These APIs are prompting a fresh rethink of the typical distribution of capabilities between servers and clients. This is likely to lead to more services and computations being offloaded to browsers, in particular within hybrid architectures. The above evolutions, away from tightly synchronized and monolithic deployments towards heterogeneous, composite and loosely coordinated distributed systems, raise a number of difficult challenges at the crossroad of theoretical distributed algorithms, system architecture, and programming frameworks. One of these challenges pertains to the growing complexity arising from these systems: as richer and more diverse services are being composed to construct whole applications, individual developers can only hope to grasp parts of the resulting systems. Similarly, weak consistency models and loose coordination mechanisms tend to lead to counter-intuitive behaviors, while only providing weak overall guarantees. This lack of systematic guarantees and understandability make it harder for practitioners to design, deploy, and validate the distributed systems they produce, leading to rising costs and high entry barriers.

In order to address these challenges, we argue that modern-day distributed systems require new principled algorithms, approaches, and architectural patterns able to provide sound foundations to their development while guaranteeing robust service guarantees, thus lowering the cost of their development and maintenance, increasing their reliability, and rendering them technically approachable to a wider audience.

2.3. Highly Personalized On-Line Services

Ever increasing volumes of data are being produced and made available from a growing number of sources (Internet of Things sensors, open data repositories, user-generated content services). As a result, digital users find it increasingly difficult to face the data deluge they are subjected to without additional help. This difficulty has fueled the rise of notification solutions over traditional search, in order to push few but relevant information items to users rather than leave them to sieve through a large mass of non-curated data. To provide such personalized services, most companies rely today on centralized or tightly coupled systems hosted in data centers or in the cloud. These systems use advanced data-mining and machine learning techniques to deliver enhanced, personalized, services to users and companies, and often exploit highly parallelized data analytics frameworks such as Spark, and Flink.

Selecting the best information for a user in order to provide a personalized experience requires however to gather enough information about this user, which raises a number of important technical challenges and privacy protection issues. More precisely, this concentration poses strong risks to the privacy of users, and limits the scope of personalization to tightly integrated datasets. The use of large monolithic infrastructures also limits the use of machine learning and personalization to situations in which data is fully available to the organization managing the underlying computing infrastructure. This set-up prevents for instance cases in which sensitive data may not be shared freely, but might be of mutual interest to several independent participants in order to construct common machine learning models usable by all. Such situations occur for instance in the context of the mining of health-records by independent health-organizations, or in the collective harnessing of individual on-line profiles for personalization purpose by private users.

Alternative decentralized approaches that eschew the need for a central all-encompassing authority holds the promise of delivering knowledge while protecting individual participants. Constructing such systems requires however to address the inherent tension between the need to limit sensitive individual leaks, while maximizing collectively gained insights. Answering this tension calls on techniques and approaches from distributed systems, information theory, security, and randomized processes, making it a rich and dense research area, with a high impact potential. The problem of distributed privacy in a digital interconnected age further touches on interdisciplinary questions of Law, Sociology and Public Policy, which we think can only be explored in collaboration with colleagues from these fields.

2.4. Social Collaboration Platforms

One-line social networks have had a fundamental and lasting impact on the Internet. In recent years, numerous applications have appeared that go beyond the services originally provided by “pure” on-line social networks, such as posting messages or maintaining on-line “friendship” links. These new applications seek to organize and coordinate users, often in the context of the sharing economy, for instance in order to facilitate car-sharing (e.g. BlaBla car, https://www.blablacar.com/), short-term renting (e.g. AirBnB, https://www.airbnb.com/), and peer-to-peer financial services (e.g. Lending Club, https://www.lendingclub.com/). Some systems, such as Bitcoin or Ethereum, have given rise to new distributed protocols combining elements of cryptography and distribution that are now largely discussed in the research community, and have attracted the attention of policy makers and leading financial actors.

The challenges faced by such social applications blend in many ways issues already discussed in the two previous subsections and cast them in an application-driven context. These social collaboration platforms require mechanisms that go beyond pure message propagation, with stricter consistency and robustness guarantees. Because they involve connected users, these applications must provide usable solutions, in particular in terms of latency and availability. At the same time, because they manipulate real-world transactions and objects (money, cars, accommodations) they must also provide a high level of consistency and guarantees. Many of these applications further operate at a planetary scale, and therefore also face stark scalability issues, that make them highly interesting case studies to investigate innovative architectures combining decentralized and centralized elements.

Formalizing and characterizing the needs and behaviors of these new applications seems particularly interesting in order to provide the fertile ground for new systems and novel theoretical work. The area of social applications also offers avenues for knowledge transfer and societal impact, along two dimensions. First, practical and usable approaches, back by a deep understanding of the foundation of distribution and coordination, are likely to find applications in future systems. Second, developers of complex social applications are often faced with a lack of robust scalable services [1] that can be easily exploited to harness the latest understanding of large-scale distributed coordination. We therefore think these applications offer an opportunity to design and deliver modular reusable bricks that can be easily appropriated by a large population of innovative developers without requiring the level of deep understanding usually necessary to implement these solutions from scratch. Providing such reusable bricks is however difficult, as many interesting formal properties are not composable, and a unified composable theory of distributed systems still need to be fully articulated.

3. Research Program

3.1. Overview

In order to progress in the four fields described above, the WIDE team is developing a research program which aims to **help developers control and master the inherent uncertainties and performance challenges brought by scale and distribution**.

More specifically, our program revolves around four key challenges.

- Challenge 1: Designing Hybrid Scalable Architectures,
- Challenge 2: Constructing Personalizable Privacy-Aware Distributed Systems,
- Challenge 3: Understanding Controllable Network Diffusion Processes,
- Challenge 4: Systemizing Modular Distributed Computability and Efficiency.

These four challenges have in common **the inherent tension between coordination and scalability in large-scale distributed systems**: strong coordination mechanisms can deliver strong guarantees (in terms of consistency, agreement, fault-tolerance, and privacy protection), but are generally extremely costly and inherently non-scalable if applied indiscriminately. By contrast, highly scalable coordination approaches (such as epidemic protocols, eventual consistency, or self-organizing overlays) perform much better when the size of a system increases, but do not, in most cases, provide any strong guarantees in terms of consistency or agreement.

[1] The repeated debugging of MongoDB's replication algorithm (e.g. see https://aphyr.com/posts/338-jepsen-mongodb-3-4-0-rc3) is a telling illustration of the difficulties encountered by development teams when building such platforms.

The above four challenges explore these tensions from *four complementary angles*: from an architectural perspective (Challenge 1), from the point of view of a fundamental system-wide guarantee (privacy protection, Challenge 2), looking at one universal scalable mechanism (network diffusion, Challenge 3), and considering the interplay between modularity and computability in large-scale systems (Challenge 4). These four challenges range from practical concerns (Challenges 1 and 2) to more theoretical questions (Challenges 3 and 4), yet present *strong synergies* and *fertile interaction points*. E.g. better understanding network diffusion (Challenge 3) is a key enabler to develop more private decentralized systems (Challenge 2), while the development of a theoretically sound modular computability hierarchy (Challenge 4) has a direct impact on our work on hybrid architectures (Challenge 1).

3.2. Hybrid Scalable Architectures

The rise of planetary-scale distributed systems calls for novel software and system architectures that can support user-facing applications while scaling to large numbers of devices, and leveraging established and emerging technologies. The members of WIDE are particularly well positioned to explore this avenue of research thanks to their experience on de-concentrated architectures combining principles from both decentralized peer-to-peer [44], [54] systems and hybrid infrastructures (i.e. architectures that combines centralized or hierarchical elements, often hosted in well-provisioned data-centers, and a decentralized part, often hosted in a peer-to-peer overlay) [48]. In the short term, we aim to explore two axes in this direction: browser-based communication, and micro services.

3.2.1. Browser-based fog computing

The dramatic increase in the amount of data being produced and processed by connected devices has led to paradigms that seek to decentralize the traditional cloud model. In 2011 Cisco [45] introduced the vision of *fog computing* that combines the cloud with resources located at the edge of the network and in between. More generally, the term *edge computing* has been associated with the idea of adding edge-of-the network storage and computation to traditional cloud infrastructures [40].

A number of efforts in this directions focus on specific hardware, e.g. fog nodes that are responsible for connected IoT devices [46]. However, many of today's applications run within web browsers or mobile phones. In this context, the recent introduction of the WebRTC API, makes it possible for browsers and smartphones to exchange directly between each other, enabling mobile, or browser-based decentralized applications. Maygh [68], for example, uses the WebRTC API to build a decentralized Content Delivery Network that runs solely on web browsers. The fact that the application is hosted completely on a web server and downloaded with enabled websites means that webmasters can adopt the Content Delivery Network (CDN) without requiring users to install any specific software.

For us, the ability of browsers to communicate with each other using the WebRTC paradigm provides a novel playground for new programming models, and for a *browser-based fog architecture* combining both a centralized, cloud-based part, and a decentralized, browser-supported part.

This model offers tremendous potential by making edge-of-the-network resources available through the interconnection of web-browsers, and offers new opportunities for the protection of the personal data of end users. But consistently engineering browser-based components requires novel tools and methodologies.

In particular, WebRTC was primarily designed for exchanging media and data between two browsers in the presence of a coordinating server. Its complex mechanisms for connection establishment make many of the existing peer-to-peer protocols inefficient. To address this challenge, we plan to consider two angles of attack. First, we plan to design novel protocols that take into account the specific requirements set by this new technology. Second, we envisage to investigate variants of the current WebRTC model with cheaper connection-establishment protocols, in order to provide lower delays and bandwidth consumption in large-scale browser-based applications.

We also plan to address the trade-offs associated with hybrid browser-cloud models. For example, when should computation be delegated to browsers and when should it be executed on the cloud in order to maximize the quality of service? Or, how can a decentralized analytics algorithms operating on browser-based data complement or exploit the knowledge built by cloud-based data analytics solutions?

3.2.2. Emergent micro-service deployment and management

Micro-services tend to produce fine-grained applications in which many small services interact in a loosely coupled manner to produce a wide range of services within an organization. Individual services need to evolve independently of each other over time without compromising the availability of the overall application. Lightweight isolation solutions such as containers (Docker, ...), and their associated tooling ecosystem (e.g. Google's Borg [67], Kubernetes [43]) have emerged to facilitate the deployment of large-scale micro-service-based applications, but only provide preliminary solutions for key concerns in these systems, which we would like to investigate and extend.

Most of today's on-line computer systems are now too large to evolve in monolithic, entirely pre-planned ways. This applies to very large data centres, for example, where the placement of virtual machines to reduce heating and power consumption can no longer be treated using top-down exhaustive optimisation approaches beyond a critical size. This is also true of social networking applications, where different mechanisms—e.g. to spread news notifications, or to recommend new contacts—must be adapted to the different sub-communities present in the system.

To cope with the inherent complexity of building complex loosely-coupled distributed systems while fostering and increasing efficiency, maintainability, and scalability, we plan to study how novel programming techniques based on declarative programming, components and epidemic protocols can help design, deploy, and maintain self-adaptive structures (e.g. placement of VM) and mechanisms (e.g. contact recommendations) that are optimized to the local context of very large distributed systems. To fulfill this vision, we plan to explore a three-pronged strategy to raise the level of programming abstraction offered to developers.

- First, we plan to explore the use of high-level domain-specific languages (DSL) to declare how large-scale topologies should be achieved, deployed, and maintained. Our vision is a declarative approach to describe how to combine, deploy and orchestrate micro-services in an abstract manner thus abstracting away developers from the underlying cloud infrastructures, and from the intricacies involved in writing low-level code to build a large-scale distributed application that scales. With this effort, we plan notably to directly support the twin properties of *emergence* (the adaptation "from within") and *differentiation* (the possibility from parts of the system to diverge while still forming a whole). Our central objective is to search for principled programming constructs to support these two capabilities using a modular and incremental software development approach.
- On a second strand of work, we plan to investigate how unikernels enable smaller footprints, more optimization options, and faster boot times for micro-services. Isolating micro-services into VMs is not the most adequate approach as it requires the use of hypervisors, or virtual machine monitors (VMMs), to virtualize hardware resources. VMMs are well known to be heavyweight with both boot and run time overheads that may have a strong impact on performances. Unikernels seem to offer the right balance between performance and flexibility to address this challenge. One of the key underlying challenges is to compile directly the aforementioned provided DSL to a dedicated and customized machine image, ready to be deployed directly on top of a large set of bare metal servers.
- Depending on the workload it is subjected to, and the state of its execution environment (network, VMs), a large-scale distributed application may present erratic or degraded performance that is hard to anticipate and plan for. There is therefore a strong need to adapt dynamically the way resources are allocated to a running application. We would like to study how the DSL approach we envisage can be extended to enable developers to express orchestration algorithms based on machine learning algorithms.

3.3. Personalizable Privacy-Aware Distributed Systems

On-line services are increasingly moving towards an in-depth analysis of user data, with the objective of providing ever better personalization. But in doing so, personalized on-line services inevitably pose risks to the privacy of users. Eliminating, or even reducing these risks raises important challenges caused by the inherent trade-off between the level of personalization users wish to achieve, and the amount of information they are willing to reveal about themselves (explicitly or through the many implicit sources of digital information such as smart homes, smart cars, and IoT environments).

At a general level, we would like to address these challenges through protocols that can provide access to unprecedented amounts of data coming from sensors, users, and documents published by users, while protecting the privacy of individuals and data sources. To this end, we plan to rely on our experience in the context of distributed systems, recommender systems, and privacy, as well as in our collaborations with experts in neighboring fields such as machine learning, and security. In particular, we aim to explore different privacy-utility tradeoffs that make it possible to provide differentiated levels of privacy guarantees depending on the context associated with data, on the users that provide the data, and on those that access it. Our research targets the general goal of privacy-preserving decentralized learning, with applications in different contexts such as user-oriented applications, and the Internet-of-Things (IoT).

3.3.1. Privacy-preserving decentralized learning

Personalization and recommendation can be seen as a specific case of general machine learning. Production-grade recommenders and personalizers typically centralize and process the available data in one location (a data-center, a cloud service). This is highly problematic, as it endangers the privacy of users, while hampering the analysis of datasets subject to privacy constraints that are held by multiple independent organizations (such as health records). A decentralized approach to machine learning appears as a promising candidate to overcome these weaknesses: if each user or participating organization keeps its data, while only exchanging gradient or model information, privacy leaks seem less likely to occur.

In some cases, decentralized learning may be achieved through relatively simple adaptations of existing centralized models, for instance by defining alternative learning models that may be more easily decentralized. But in all cases, processing growing amounts of information calls for high-performance algorithms and middleware that can handle diverse storage and computation resources, in the presence of dynamic and privacy-sensitive data. To reach this objective, we will therefore leverage our work in distributed and privacy-preserving algorithms and middleware [47], [49], [50] as well as the results of our work on large-scale hybrid architectures in Objective 1.

3.3.2. Personalization in user-oriented applications

As a first application perspective, we plan to design tools that exploit decentralized analytics to enhance user-centric personalized applications. As we observed above, such applications exhibit an inherent trade-off between personalization quality and privacy preservation. The most obvious goal in this direction consists in designing algorithms that can achieve high levels of personalization while protecting sensitive user information. But an equally important one consists in personalizing the trade-off itself by adapting the quality of the personalization provided to a user to his/her willingness to expose information. This, like other desirable behaviors, appears at odds with the way current systems work. For example, a user of a recommender system that does not reveal his/her profile information penalizes other users causing them to receive less accurate recommendations. We would like to mitigate this situation by means of protocols that reward users for sharing information. On the one hand, we plan to take inspiration from protocols for free-riding avoidance in peer-to-peer systems [51], [56]. On the other hand, we will consider blockchains as a tool for tracking and rewarding data contributions. Ultimately, we aim at enabling users to configure the level of privacy and personalization they wish to experience.

3.3.3. Privacy preserving decentralized aggregation

As a second setting we would like to consider target applications running on constrained devices like in the Internet-of-Things (IoT). This setting makes it particularly important to operate on decentralized data in a light-weight privacy-preserving manner, and further highlights the synergy between this objective and Objective 1. For example, we plan to provide data subjects with the possibility to store and manage their data locally on their own devices, without having to rely on third-party managers or aggregators, but possibly storing less private information or results in the cloud. Using this strategy, we intend to design protocols that enable users themselves, or third-party companies to query distributed data in aggregate form, or to run data analytics processes on a distributed set of data repositories, thereby gathering knowledge without violating the privacy of other users. For example, we have started working on the problem of computing an aggregate function over a subset of the data in a distributed setting. This involves two major steps: selection and aggregation. With respect to selection, we envision defining a decentralized data-selection operation that can apply a selection predicate without violating privacy constraints. With respect to aggregation, we will continue our investigation of lightweight protocols that can provide privacy with limited computational complexity [41].

3.4. Network Diffusion Processes

Social, biological, and technological networks can serve as conduits for the spread of ideas, trends, diseases, or viruses. In social networks, rumors, trends and behaviors, or the adoption of new products, spread from person to person. In biological networks, diseases spread through contact between individuals, and mutations spread from an individual to its offsprings. In technological networks, such as the Internet and the power grid, viruses and worms spread from computer to computer, and power failures often lead to cascading failures. The common theme in all the examples above is that the rumor, disease, or failure starts out with a single or a few individual nodes, and propagates through the network, from node to node, to reach a potentially much larger number of nodes.

These types of *network diffusion processes* have long been a topic of study in various disciplines, including sociology, biology, physics, mathematics, and more recently, computer science. A main goal has been to devise mathematical models for these processes, describing how the state of an individual node can change as a function of the state of its neighbors in the network, and then analyse the role of the network structure in the outcome of the process. Based on our previous work, we would like to study to what extent one can affect the outcome of the diffusion process by controlling a small, possibly carefully selected fraction of the network.

For example, we plan to explore how we may increase the spread or speed of diffusion by choosing an appropriate set of seed nodes (a standard goal in viral marketing by word-of-mouth), or achieve the opposite effect either by choosing a small set of nodes to remove (a goal in immunization against diseases), or by seeding a competing diffusion (e.g., to limit the spread of misinformation in a social network).

Our goal is to provide a framework for a systematic and rigorous study of these problems. We will consider several standard diffusion models and extensions of them, including models from mathematical sociology, mathematical epidemiology, and interacting particle systems. We will consider existing and new variants of spread maximization/limitation problems, and will provide (approximation) algorithms or show negative (inapproximability) results. In case of negative results, we will investigate general conditions that make the problem tractable. We will consider both general network topologies and specific network models, and will relate the efficiency of solutions to structural properties of the topology. Finally, we will use these insights to engineer new network diffusion processes for efficient data dissemination.

3.4.1. Spread maximization

Our goal is in particular to study spread maximization in a broader class of diffusion processes than the basic independent cascade (IC) and linear threshold (LT) models of influence [60], [61], [62] that have been studied in this context so far. This includes the *randomized rumor spreading (RS)* model for information dissemination [53], *biased* versions of the *voter model* [57] modelling influence, and the (graph-based) *Moran processes* [64] modelling the spread of mutations. We would like to consider several natural versions of the spread maximization problem, and the relationships between them. For these problems we will use the greedy algorithm and the submodularity-based analytical framework of [60], and will also explore new approaches.

3.4.2. Immunization optimization

Conversely we would also like to explore immunization optimization problems. Existing works on these types of problem assume a *perfect-contagion* model, i.e., once a node gets infected, it deterministically infects all its non-immunized neighbors. We plan to consider various diffusion processes, including the standard *susceptible–infected* (SI), *susceptible–infected–recovered* (SIR) and *susceptible–infected–susceptible* (SIS) epidemic models, and explore the extent to which results and techniques for the perfect-contagion model carry over to these probabilistic models. We will also investigate whether techniques for spread maximization could be applied to immunization problems.

Some immunization problems are known to be hard to approximate in general graphs, even for the perfect-contagion model, e.g., the fixed-budget version of the fire-fighter problem cannot be approximated to any $n^{1-\epsilon}$ factor [42]. This strand of work will consider restricted graph families, such as trees or graphs of small treewidth, for such problems. In addition, for some immunization problems, there is a large gap between the best known approximation algorithm and the best known inaproximability result, and we would like to make progress in reducing these gaps.

3.5. Systemizing Modular Distributed Computability and Efficiency

The applications and services envisaged in Objectives 1 and 2 will lead to increasingly complex and multifaceted systems. Constructing these novel hybrid and decentralized systems will naturally push our need to understand distributed computing beyond the current state of the art. These trends therefore demand research efforts in establishing sound theoretical foundations to allow everyday developers to master the design, properties and implementation of these systems. We plan to investigate these foundations along two directions: first by studying novel approaches to some fundamental problems of *mutual exclusion and distributed coordination*, and second by exploring how we can build a *comprehensive and modular framework* capturing the foundations of *distributed computation*.

3.5.1. Randomized algorithm for mutual exclusion and coordination

To exploit the power of massive distributed applications and systems (such as those envisaged in Objectives 1 and 2) or multiple processors, algorithms must cope with the scale and asynchrony of these systems, and their inherent instability, e.g., due to node, link, or processor failures. Our goal is to explore the power and limits of randomized algorithms for large-scale networks of distributed systems, and for shared memory multi-processor systems, in effect providing fundamental building blocks to the work envisioned in Objectives 1 and 2.

For shared memory systems, randomized algorithms have notably proved extremely useful to deal with asynchrony and failures. Sometimes probabilistic algorithms provide the only solution to a problem; sometimes they are more efficient; sometimes they are simply easier to implement. We plan to devise efficient algorithms for some of the fundamental problems of shared memory computing, such as mutual exclusion, renaming, and consensus.

In particular, looking at the problem of *mutual exclusion*, it is desirable that mutual exclusion algorithms be *abortable*. This means that a process that is trying to lock the resource can abort its attempt in case it has to wait too long. Abortability is difficult to achieve for mutual exclusion algorithms. We will try to extend our algorithms for the *cache-coherent* (CC) and the *distributed shared memory* (DSM) model in order to make them abortable, while maintaining expected constant *Remote Memory References* (RMRs) complexity, under optimistic system assumptions. In order to achieve this, the algorithm will use strong synchronization primitives, called compare-and-swap objects. As part of our collaboration with the University of Calgary, we will work on implementing those objects from registers in such a way that they also allow aborts. Our goal is to build on existing non-abortable implementations [55]. We plan then later to use these objects as building blocks in our mutual exclusion algorithm, in order to make them work even if the system does not readily provide such primitives.

We have also started working on blockchains, as these represent a new and interesting trade-off between probabilistic guarantees, scalability, and system dynamics, while revisiting some of the fundamental questions and limitations of consensus in fault-prone asynchronous systems.

3.5.2. Modular theory of distributed computing

Practitioners and engineers have proposed a number of reusable frameworks and services to implement specific distributed services (from Remote Procedure Calls with Java RMI or SOAP-RPC, to JGroups for group communication, and Apache Zookeeper for state machine replication). In spite of the high conceptual and practical interest of such frameworks, many of these efforts lack a sound grounding in distributed computation theory (with the notable exceptions of JGroups and Zookeeper), and often provide punctual and partial solutions for a narrow range of services. We argue that this is because we still lack a generic framework that unifies the large body of fundamental knowledge on distributed computation that has been acquired over the last 40 years.

To overcome this gap we would like to develop a systematic model of distributed computation that organizes the functionalities of a distributed computing system into reusable modular constructs assembled via well-defined mechanisms that maintain sound theoretical guarantees on the resulting system. This research vision arises from the strong belief that distributed computing is now mature enough to resolve the tension between the social needs for distributed computing systems, and the lack of a fundamentally sound and systematic way to realize these systems.

To progress on this vision, we plan in the near future to investigate, from a distributed software point of view, the impact due to failures and asynchrony on the layered architecture of distributed computing systems. A first step in this direction will address the notions of *message adversaries* (introduced a long time ago in [66]) and *process adversaries* (investigated in several papers, e.g. [65], [52], [58], [59], [63]). The aim of these notions is to consider failures, not as "bad events", but as part of the normal behavior of a system. As an example, when considering round-based algorithms, a message adversary is a daemon which, at every round, is allowed to suppress some messages. The aim is then, given a problem P, to find the strongest adversary under which P can be solved ("strongest" means here that giving more power to the adversary makes the problem impossible to solve). This work will allow us to progress in terms of general *layered* theory of distributed computing, and allow us to better *map* distributed computing models and their relations, in the steps of noticeable early efforts in this direction [65], [39].

4. Highlights of the Year

4.1. Highlights of the Year

Awards

Michel Raynal is the recipient of the 2018 IEEE award for Outstanding Technical Achievement in Distributed Computing.

5. New Software and Platforms

5.1. WebGC

Web-based Gossip Communication
KEYWORDS: Epidemic protocols - Gossip protocols - Peer-to-peer - Web - Personalized systems - Decentralized architectures - Recommendation systems - WebRTC - Decentralized web
SCIENTIFIC DESCRIPTION: The library currently includes the implementation of two peer sampling protocols, Cyclon and the generic peer-sampling protocol from, as well as a clustering protocol. All protocols implement a common GossipProtocol "interface"

FUNCTIONAL DESCRIPTION: WebGC consists of a WebRTC-based library that supports gossip-based communication between web browsers and enables them to operate with Node-JS applications. WebGC comprises the implementation of standard gossip protocols such as Peer Sampling or Clustering, and simplifies the development of new protocols. It comprises a decentralized signaling service that makes it easier to build completely decentralized browser-based applications.

- Participants: Anne-Marie Kermarrec, Davide Frey, Matthieu Simonin and Raziel Carvajal Gomez
- Contact: Davide Frey

5.2. YALPS

KEYWORDS: Traffic-shaping - Nat traversal - Experimentation - Peer-to-peer - Simulator - Deployment
FUNCTIONAL DESCRIPTION: YALPS is an open-source Java library designed to facilitate the development, deployment, and testing of distributed applications. Applications written using YALPS can be run both in simulation and in real-world mode without changing a line of code or even recompiling the sources. A simple change in a configuration file will load the application in the proper environment. A number of features make YALPS useful both for the design and evaluation of research prototypes and for the development of applications to be released to the public. Specifically, YALPS makes it possible to run the same application as a simulation or in a real deployment. Applications communicate by means of application-defined messages which are then routed either through UDP/TCP or through YALPS's simulation infrastructure. In both cases, YALPS's communication layer offers features for testing and evaluating distributed protocols and applications. Communication channels can be tuned to incorporate message losses or to constrain their outgoing bandwidth. Finally, YALPS includes facilities to support operation in the presence of NATs and firewalls using relaying and NAT-traversal techniques. The implementation of YALPS includes approximately 16K lines of code, and is used in several projects by ASAP, including HEAP, AllYours-P2P, and Behave.

- Participants: Anne Marie Kermarrec, Arnaud Jegou, Davide Frey, Heverson Borba Ribeiro and Maxime Monod
- Contact: Davide Frey
- URL: http://yalps.gforge.inria.fr/

5.3. KIFF

KIFF: An impressively fast and efficient JAVA library for KNN construction
KEYWORD: KNN
FUNCTIONAL DESCRIPTION: This package implements the KIFF algorithm reported in [1]. KIFF is a generic, fast and scalable K-Nearest-Neighbor graph construction algorithm. This algorithm connects each object to its k most similar counterparts, according to a given similarity metric. In term of comparison, this package implements also HYREC [2] and NN-DESCENT [3]. The standalone program implements cosine similarity only, however this library supports arbitrary similarity measures.

[1] Antoine Boutet, Anne-Marie Kermarrec, Nupur Mittal, Francois Taiani. Being prepared in a sparse world: the case of KNN graph construction. ICDE 2016, Finland.

- Partner: LIRIS
- Contact: Antoine Boutet

6. New Results

6.1. Scalable Systems

6.1.1. Nobody cares if you liked Star Wars: KNN graph construction on the cheap.

Participants: Olivier Ruas, François Taïani.

K-Nearest-Neighbors (KNN) graphs play a key role in a large range of applications. A KNN graph typically connects entities characterized by a set of features so that each entity becomes linked to its k most similar counterparts according to some similarity function. As datasets grow, KNN graphs are unfortunately becoming increasingly costly to construct, and the general approach, which consists in reducing the number of comparisons between entities, seems to have reached its full potential. In work [27] we propose to overcome this limit with a simple yet powerful strategy that samples the set of features of each entity and only keeps the least popular features. We show that this strategy outperforms other more straightforward policies on a range of four representative datasets: for instance, keeping the 25 least popular items reduces computational time by up to 63%, while producing a KNN graph close to the ideal one.

This work was done in collaboration with Anne-Marie Kermarrec (Mediego/EPFL).

6.1.2. Pleiades: Distributed structural invariants at scale

Participants: Simon Bouget, David Bromberg, Adrien Luxey, François Taïani.

Modern large scale distributed systems increasingly espouse sophisticated distributed architectures characterized by complex distributed structural invariants. Unfortunately, maintaining these structural invariants at scale is time consuming and error prone, as developers must take into account asynchronous failures, loosely coordinated sub-systems and network delays. To address this problem, we propose Pleiades [31], a new framework to construct and enforce large-scale distributed structural invariants under aggressive conditions. Pleiades combines the resilience of self-organizing overlays, with the expressiveness of an assembly-based design strategy. The result is a highly survivable framework that is able to dynamically maintain arbitrary complex distributed structures under aggressive crash failures. Our evaluation shows in particular that Pleiades is able to restore the overall structure of a 25,600 node system in less than 11 asynchronous rounds after half of the nodes have crashed.

6.1.3. CASCADE: Reliable distributed session handoff for continuous interaction across devices

Participants: David Bromberg, Adrien Luxey, François Taïani.

Allowing users to navigate seamlessly between their personal devices while protecting their privacy remains today an ongoing challenge. Existing solutions rely on peer-to-peer designs, and blindly flood the network with session messages. It is particularly hard to come up with proposals that are both cost-efficient and dependable while relying on poorly connected mobile appliances. In [24] we propose Cascade, a distributed protocol to share applicative sessions among one's devices. Our proactive session handoff algorithm takes inspiration from the BitTorrent P2P file sharing protocol, but adapts it to the specific characteristics of our problem. It eschews in particular trackers, and limits the seeders of each session to the devices most likely to be used next, as computed by a decentralized aggregation protocol. A key aspect of our approach is to trade off network costs for reliability, while providing a faster session handoff than centralized solutions in the vast majority of the cases.

6.1.4. Sprinkler: A probabilistic dissemination protocol to provide fluid user interaction in multi-device ecosystems

Participants: David Bromberg, Adrien Luxey, François Taïani.

Offering fluid multi-device interactions to users while protecting their privacy largely remains an ongoing challenge. Existing approaches typically use a peer-to-peer design and flood session information over the network, resulting in costly and often unpractical solutions. In [29], we propose Sprinkler, a decentralized probabilistic dissemination protocol that uses a gossip-based learning algorithm to intelligently propagate session information to devices a user is most likely to use next. Our solution allows designers to efficiently trade off network costs for fluidity, and is for instance able to reduce network costs by up to 80% against a flooding strategy while maintaining a fluid user experience.

This work was done in collaboration with Fabio Costa, Ricardo Da Rocha and Vinicius Lima from the Universidade Federal de Goias (UFG).

6.2. Personalization and Privacy

6.2.1. GoldFinger

Participants: Olivier Ruas, François Taïani.

In work [37] we propose fingerprinting, a new technique that consists in constructing compact, fast-to-compute and privacy-preserving representation of datasets. We illustrate the effectiveness of our approach on the emblematic big data problem of K-Nearest-Neighbor (KNN) graph construction and show that fingerprinting can drastically accelerate a large range of existing KNN algorithms, while efficiently obfuscating the original data, with little to no overhead. Our extensive evaluation of the resulting approach (dubbed GoldFinger) on several realistic datasets shows that our approach delivers speed-ups up to 78.9% compared to the use of raw data while only incurring a negligible to moderate loss in terms of KNN quality. To convey the practical value of such a scheme, we apply it to item recommendation, and show that the loss in recommendation quality is negligible.

This work was done in collaboration with Rachid Guerraoui (EPFL) and Anne-Marie Kermarrec (Mediego/EPFL).

6.2.2. Collaborative filtering under a Sybil attack: Similarity metrics do matter!

Participant: Davide Frey.

Recommendation systems help users identify interesting content, but they also open new privacy threats. For this reason, in [22] we deeply analyzed the effect of a Sybil attack that tries to infer information on users from a user-based collaborative-filtering recommendation systems. We evaluated the impact of different similarity metrics used to identity users with similar tastes in the trade-off between recommendation quality and privacy. Based on our results, we proposed and evaluated a novel similarity metric that combines the best of both worlds: a high recommendation quality with a low prediction accuracy for the attacker. Our experiments, on a state-of-the-art recommendation framework and on real datasets showed that existing similarity metrics exhibit a wide range of behaviors in the presence of Sybil attacks, while our new similarity metric consistently achieves the best trade-off while outperforming state-of-the-art solutions.

This work was carried out in collaboration with Antoine Boutet from INSA Lyon, former-intern Florestan De Moor, Rachid Guerraoui and Antoine Rault from EPFL, and Anne-Marie Kermarrec from Mediego.

6.3. Network and Graph Algorithms

6.3.1. Rumor spreading and conductance

Participant: George Giakkoupis.

In [16], we study the completion time of the PUSH-PULL variant of rumor spreading, also known as randomized broadcast. We show that if a network has n nodes and conductance ϕ then, with high probability, PUSH-PULL will deliver the message to all nodes in the graph within $O(\log n/\phi)$ many communication rounds. This bound is best possible. We also give an alternative proof that the completion time of PUSH-PULL is bounded by a polynomial in $\log n/\phi$, based on graph sparsification. Although the resulting asymptotic bound is not optimal, this proof shows an interesting and, at the outset, unexpected connection between rumor spreading and graph sparsification. Finally, we show that if the degrees of the two endpoints of each edge in the network differ by at most a constant factor, then both PUSH and PULL alone attain the optimal completion time of $O(\log n/\phi)$, with high probability.

This work was done in collaboration with Flavio Chierichetti (Sapienza University of Rome), Silvio Lattanzi (Google Research), and Alessandro Panconesi (Sapienza University of Rome).

6.3.2. Tight bounds for coalescing-branching random walks on regular graphs

Participant: George Giakkoupis.

A Coalescing-Branching Random Walk (CoBra) is a natural extension to the standard random walk on a graph. The process starts with one pebble at an arbitrary node. In each round of the process every pebble splits into k pebbles, which are sent to k random neighbors. At the end of the round all pebbles at the same node coalesce into a single pebble. The process is also similar to randomized rumor spreading, with each informed node pushing the rumor to k random neighbors each time it receives a copy of the rumor. Besides its mathematical interest, this process is relevant as an information dissemination primitive and a basic model for the spread of epidemics. In [21], we study the cover time of CoBra walks, which is the time until each node has seen at least one pebble. Our main result is a bound of $O(\log n/\phi)$ rounds with high probability on the cover time of a CoBra walk with $k = 2$ on any regular graph with n nodes and conductance ϕ. This bound improves upon all previous bounds in terms of graph expansion parameters. Moreover, we show that for any connected regular graph the cover time is $O(n \log n)$ with high probability, independently of the expansion. Both bounds are asymptotically tight. Since our bounds coincide with the worst-case time bounds for Push rumor spreading on regular graphs until all nodes are informed, this raises the question whether CoBra walks and Push rumor spreading perform similarly in general. We answer this negatively by separating the cover time of CoBra walks and the rumor spreading time of Push by a super-polylogarithmic factor on a family of tree-like regular graphs.

This work was done in collaboration with Petra Berenbrink and Peter Kling from the University of Hamburg.

6.3.3. The quadratic shortest path problem: Complexity, approximability, and solution methods

Participant: Davide Frey.

In work [20] we considered the problem of finding a shortest path in a directed graph with a quadratic objective function (the QSPP). We show that the QSPP cannot be approximated unless $\mathsf{P} = \mathsf{NP}$. For the case of a convex objective function, we presented an n-approximation algorithm, where n is the number of nodes in the graph, and we proved APX-hardness. Furthermore, we proved that even if only adjacent arcs play a part in the quadratic objective function, the problem still cannot be approximated unless $\mathsf{P} = \mathsf{NP}$. In order to solve the general problem we first proposed a mixed integer programming formulation, and then devised an efficient exact Branch-and-Bound algorithm for the general QSPP. This algorithm computes lower bounds by considering a reformulation scheme that is solvable through a number of minimum-cost-flow problems. We carried out computational experiments solving to optimality different classes of instances with up to 1000 nodes.

This work was carried out in collaboration with Borzou Rostami from Polytechnique Montréal, Adreé Chasssein and Michael Hopf from TU Kaiserslautern, Christoph Buchheim from TU Dortmund, Federico Malucelli from Politecnico di Milano, and Marc Goerigk from Lancaster University.

6.3.4. Weighting past on the geo-aware state deployment problem

Participant: François Taïani.

The geographical barrier between mobile devices and mobile application servers (typically hosted in the Cloud) imposes an unavoidable latency and jitter that negatively impacts the performance of modern mobile systems. Fog Computing architectures can mitigate this impact if there is a middleware service able to correctly partition and deploy the state of an application at optimal locations. Geo-aware state deployment is challenging as it must consider the mobility of the devices and the dependencies arising when multiple devices concurrently manipulate the same application state. In [28], we propose a range of new object-graph-based strategies for geo-aware state deployment. In particular, our investigation focuses on understanding the role of preserving previously observed associations between state items on application performance.

This work was performed in collaboration with Diogo Lima and Hugo Miranda from the University of Lisbon (Portugal).

6.3.5. Mind the gap: Autonomous detection of partitioned MANET systems using opportunistic aggregation

Participants: Simon Bouget, David Bromberg, François Taïani.

Mobile Ad-hoc Networks (MANETs) use limited-range wireless communications and are thus exposed to partitions when nodes fail or move out of reach of each other. Detecting partitions in MANETs is unfortunately a nontrivial task due to their inherently decentralized design and limited resources such as power or bandwidth. In [32], we propose a novel and fully decentralized approach to detect partitions (and other *large* membership changes) in MANETs that is both accurate and resource efficient. We monitor the current composition of a MANET using the lightweight aggregation of compact membership-encoding filters. Changes in these filters allow us to infer the likelihood of a partition with a quantifiable level of confidence. We first present an analysis of our approach, and show that it can detect close to 100% of partitions under realistic settings, while at the same time being robust to false positives due to churn or dropped packets. We perform a series of simulations that compare against alternative approaches and confirm our theoretical results, including above 90% accurate detection even under a 40% message loss rate.

This work was performed in collaboration with Etienne Rivière from UC Louvain (Belgium) and Hugues Mercier from University of Neuchâtel (Switzerland).

6.4. Theory of Distributed Systems

6.4.1. An improved bound for random binary search trees with concurrent insertions

Participant: George Giakkoupis.

Recently, Aspnes and Ruppert (DISC 2016) defined the following simple random experiment to determine the impact of concurrency on the performance of binary search trees: n randomly permuted keys arrive one at a time. When a new key arrives, it is first placed into a buffer of size c. Whenever the buffer is full, or when all keys have arrived, an adversary chooses one key from the buffer and inserts it into the binary search tree. The ability of the adversary to choose the next key to insert among c buffered keys, models a distributed system, where up to c processes try to insert keys concurrently. Aspnes and Ruppert showed that the expected average depth of nodes in the resulting tree is $O(\log n + c)$ for a comparison-based adversary, which can only take the relative order of arrived keys into account. In work [25], we generalize and strengthen this result. In particular, we allow an adversary that knows the actual values of all keys that have arrived, and show that the resulting expected average node depth is $D_{avg}(n) + O(c)$, where $D_{avg}(n) = 2ln(n) - Theta(1)$ is the expected average node depth of a random tree obtained in the standard unbuffered version of this experiment. Extending the bound by Aspnes and Ruppert to this stronger adversary model answers one of their open questions.

This work was done in collaboration with Philipp Woelfel (University of Calgary).

6.4.2. Acyclic strategy for silent self-stabilization in spanning forests

Participant: Anaïs Durand.

Self-stabilization is a general paradigm to enable the design of distributed systems tolerating any finite number of transient faults. Many self-stabilizing algorithms are designed using the same patterns. In [30] we formalize some of those design patterns to obtain general statements regarding both correctness and time complexity. Precisely, we study a class of algorithms devoted to networks endowed with a sense of direction describing a spanning forest whose characterization is a simple (i.e., quasi-syntactic) condition. We show that any algorithm of this class is (1) silent and self-stabilizing under the distributed unfair daemon (the weakest scheduling assumption in the considered model), and (2) has a stabilization time polynomial in moves and asymptotically optimal in rounds. Our condition mainly uses the concept of acyclic strategy, which is based on the notions of top-down and bottom-up actions. We have combined this formalization together with a notion of acyclic causality between actions and a last criteria called correct-alone (n.b., only this criteria is not syntactic) to obtain the notion of acyclic strategy. We show that any algorithm following an acyclic strategy reaches a terminal configuration in a polynomial number of moves, assuming a distributed unfair daemon. Hence, if its terminal configurations satisfy the specification, the algorithm is both silent and self-stabilizing. Unfortunately, we show that this condition is not sufficient to obtain an asymptotically optimal stabilization time in rounds. So, we enforce the acyclic strategy with the property of local mutual exclusivity to have an asymptotically optimal round complexity. We also propose a simple method to make any algorithm, that follows an acyclic strategy, locally mutually exclusive. This method has no overhead in moves. Finally, to show the versatility of our approach, we review works where our results apply.

This work was done in collaboration with Karine Altisen and Stéphane Devismes (VERIMAG, Université Grenoble Alpes).

6.4.3. Set agreement and renaming in the presence of contention-related crash failures

Participants: Anaïs Durand, Michel Raynal.

Given a predefined contention threshold λ, consider all executions in which process crashes are restricted to occur only when process contention is smaller than or equal to λ. If crashes occur after contention bypassed λ, there are no correctness guarantees (e.g., termination is not guaranteed). It is known that, when $\lambda = n{-}1$, consensus can be solved in an n-process asynchronous read/write system despite the crash of one process, thereby circumventing the well-known FLP impossibility result. Furthermore, it was shown that when $\lambda = n{-}k$ and $k \geq 2$, k-set agreement can be solved despite the crash of $2k{-}2$ processes.

In work [33] we consider two types of process crash failures: λ-constrained crash failures (as previously defined), and classical crash failures (that we call *any time* failures). We present two algorithms suited to these types of failures. The first algorithm solves k-set agreement, where $k = m + f$, in the presence of $t = 2m + f{-}1$ crash failures, $2m$ of them being $(n{-}k)$-constrained failures, and $(f{-}1)$ being any time failures. The second algorithm solves $(n + f)$-renaming in the presence of $t = m + f$ crash failures, m of them being $(n{-}t{-}1)$-constrained failures, and f being any time failures. It follows that the differentiation between λ-constrained crash failures and any time crash failures enlarges the space of executions in which the impossibility of k-set agreement and renaming in the presence of asynchrony and process crashes can be circumvented. In addition to its behavioral properties, both algorithms have a noteworthy first class property, namely, their simplicity.

This work was done in collaboration with Gadi Taubenfeld (IDC Herzliya).

6.4.4. Anonymous obstruction-free (n, k)-set agreement with $n{-}k + 1$ atomic read/write registers

Participant: Michel Raynal.

The k-set agreement problem is a generalization of the consensus problem. Namely, assuming that each process proposes a value, every non-faulty process must decide one of the proposed values, under the constraint that at most k different values are decided. This is a hard problem in the sense that it cannot be solved in a pure read/write asynchronous system, in which k or more processes may crash. One way to sidestep this impossibility result consists in weakening the termination property, requiring only that a process decides if it executes alone during a long enough period of time. This is the well-known obstruction-freedom progress condition. Consider a system of n anonymous asynchronous processes that communicate through atomic read/write registers, and such that any number of them may crash. In work [14] we address and solve the challenging open problem of designing an obstruction-free k-set agreement algorithm with only $(n{-}k + 1)$ atomic registers. From a shared memory cost point of view, our algorithm is the best algorithm known to date, thereby establishing a new upper bound on the number of registers needed to solve this problem. For the consensus case ($k = 1$), the proposed algorithm is up to an additive factor of 1 close to the best known lower bound. Further, the paper extends this algorithm to obtain an x-obstruction-free solution to the k-set agreement problem that employs $(n{-}k + x)$ atomic registers, as well as a space-optimal solution for the repeated version of k-set agreement. Using this last extension, we prove that n registers are enough for every colorless task that is obstruction-free solvable with identifiers and any number of registers.

This work was done in collaboration with Zohir Bouzid and Pierre Sutra (CNRS).

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. CIFRE Technicolor: Distributed troubleshooting of edge-compute functions (2018-2021)

Participant: François Taïani.

This project seeks to explore how recent generations of end-user gateways (or more generally end-user devices) could implement an edge-compute paradigm powered by user-side micro-services. Our vision is that the devices distributed among the homes of end-users will expose (as a service) their computing power and their ability to quickly deploy compute functions in an execution environment. In order for service and application providers to actually use the system and deploy applications, the system must however ensure an appropriate level of reliability, while simultaneously requiring a very low level of maintenance in order to address the typical size and economics of gateway deployments (at least a few tens of million units). Providing a good level of reliability in such a large system at a reasonable cost is unfortunately difficult. To address this challenge, we aim in this thesis to exploit the *natural distribution* of such large-scale user-side device deployments to quickly pinpoint problems and troubleshoot applications experiencing performance degradations.

7.2. Bilateral Grants with Industry

7.2.1. Google Focussed Grant Web Alter Ego (2013-2018)

Participants: George Giakkoupis, François Taïani.

This project addresses the problem of extracting the alter-egos of a Web user, namely profiles of like-minded users who share similar interests, across various Internet applications. The project, in collaboration with the team of Rachid Guerraoui at EPFL, runs until August 2018 and funds the PhD of Olivier Ruas, who is co-supervised by François Taïani and Anne-Marie Kermarrec.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. Web of Browser's (Brittany Region and Labex CominLabs 2019-2020)

Participant: François Taïani.

Browsers are de facto the most widely deployed execution environments in the world. Initially simple HTML readers, they now run complex applications interacting with humans and web services. The recent introduction of WebRTC has further extended the capability of browsers by introducing support for browser-to-browser communication. This turns browsers into a decentralized execution environment where interactions between human and web services are enabled without third party.

The Web of browsers is a vision where the web is serverless, ephemeral and massively decentralized. Web where pages are hosted by networks of browsers connected through WebRTC. The objective of the project is to build and experiment the Web of Browsers.

8.2. National Initiatives

8.2.1. Labex CominLab Descent (2013-2018)

Participants: Davide Frey, Michel Raynal, François Taïani.

This project (2013-2018), which also involves researchers from Nantes (LS2N, former LINA), aims to provide fundamental programming blocks to support the construction of federations of plug computers (e.g. Raspberry pi). The project's overarching vision is that everyone should be able create cheap nano-clusters of domestic servers, host data and services and federate these resources with their friends, colleagues, and families.

8.2.2. ANR Project PAMELA (2016-2020)

Participants: Davide Frey, George Giakkoupis, François Taïani.

PAMELA is a collaborative ANR project involving Inria/IRISA, Inria Lille (MAGNET team), UMPC, Mediego and Snips. The project aims at developing machine learning theories and algorithms in order to learn local and personalized models from data distributed over networked infrastructures. This project seeks to provide fundamental answers to modern information systems built by interconnecting many personal devices holding private user data in the search of personalized suggestions and recommendations. A significant asset of the project is the quality of its industrial partners, Snips and Mediego, who bring in their expertise in privacy protection and distributed computing as well as use cases and datasets.

8.2.3. ANR Project OBrowser (2016-2020)

Participants: David Bromberg, Davide Frey, François Taïani.

OBrowser is a collaborative ANR project involving Inria, the University of Nantes, the University of South Brittany, and Orange. The project emerges from the vision of designing and deploying distributed applications on millions of machines using web-enabled technologies without relying on a cloud or a central authority. OBrowser proposes to build collaborative applications through a decentralized execution environment composed of users' browsers that autonomously manages issues such as communication, naming, heterogeneity, and scalability.

8.2.4. ANR Project DESCARTES (2016-2020)

Participants: George Giakkoupis, Michel Raynal, François Taïani.

DESCARTES is a collaborative ANR project involving Inria/IRISA, Labri (U. Bordeaux), IRIF (U. Paris Diderot), Inria Paris (GANG Team), Vérimag (Grenoble), LIF (Marseilles), and LS2N (former LINA, Nantes). The DESCARTES project aims at bridging the lack of a generic theoretical framework in order to unify the large body of fundamental knowledge on distributed computation that has been acquired over the last 40 years. In particular, the project's objective is to develop a systematic model of distributed computation that organizes the functionalities of a distributed computing system into reusable modular constructs assembled via well-defined mechanisms that maintain sound theoretical guarantees on the resulting system.

8.2.5. ANR-ERC Tremplin Project NDFUSION (2016-2018)

Participant: George Giakkoupis.

NDFUSION is an 18-month ANR project awarded to George Giakkoupis to support his preparation for his upcoming ERC grant application.

The idea of intervening in a network diffusion process to enhance or retard its spread has been studied in various contexts, e.g., to increase the spread or speed of diffusion by choosing an appropriate set of seed nodes, or achieve the opposite effect either by choosing a small set of nodes to remove, or by seeding a competing diffusion (e.g., to limit the spread of misinformation in a social network).

8.2.6. Labex CominLab PROFILE (2016-2019)

Participants: David Bromberg, Davide Frey, François Taïani.

The PROFILE (2016-2019) project brings together experts from law, computer science (the Inria teams DIVERSE and ASAP/WIDE, the IRISA team DRUID) and sociology to address the challenges raised by online profiling, following a multidisciplinary approach. More precisely, the project will pursue two complementary and mutually informed lines of research: first, the project will investigate, design, and introduce a new right of opposition into privacy Law to better regulate profiling and to modify the behavior of commercial companies. Second, the project aims to provide users with the technical means they need to detect stealthy profiling techniques, and to control the extent of the digital traces they routinely produce.

8.3. International Initiatives

8.3.1. Inria Associate Teams Not Involved in an Inria International Lab

8.3.1.1. LiDiCo

Title: Aux limites du calcul réparti

International Partner (Institution - Laboratory - Researcher):

UNAM (Mexico) - Instituto de Matematicas - Sergio Rajsbaum

Start year: 2017

See also: https://sites.google.com/site/lidicoequipeassociee/

Today distributed applications are pervasive, some very successful (e.g., Internet, P2P, social networks, cloud computing), and benefit everyone, but the design and the implementation of many of them still rely on ad-hoc techniques instead of on a solid theory. The next generation of distributed applications and services will be more and more complex and demands research efforts in establishing sound theoretical foundations to be able to master their design, their properties and their implementation. This proposal is a step in this inescapable direction.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

Fábio Moreira Costa (Institute of Informatics, Federal University of Goiás, Goiânia, GO, Brazil); 17–28 September, 5–11 November, and 3–14 December 2018.

Paulo Ferreira (INESC ID Lisboa, Portugal); Associate Professor, in Sabbatical, 5 February–17 March 2018.

Hayk Saribekyan (University of Cambridge, UK); 2–12 April 2018

Sergio Rajsbaum (Instituto de Matematicas, Universidad Nacional Autonoma de Mexico, UNAM); 20 April 2018.

Thomas Sauerwald (University of Cambridge, UK); 2–12 April 2018.

8.4.2. Visits to International Teams

8.4.2.1. Research Stays Abroad

Quentin Dufour visited Etienne Rivière, Université de Louvain La Neuve UCL, 5–30 November 2018.

David Bromberg visited University of Sao Paulo, Brazil, 8–24 July 2018.

Michel Raynal visited Jiannong Cao, HK Politechnic University, 17 November–22 December 2018.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

François Taïani served as Co-Organizer of the 8th Inria/Technicolor Workshop On Systems (WOS8), Rennes, France, December 2018.

François Taïani served as Communication Chair of the 19th ACM/IFIP/Usenix International Conference on Middleware (Middleware), Rennes, France, December 2018.

9.1.2. Scientific Events Selection

9.1.2.1. Chair of Conference Program Committees

François Taïani served as PC Co-Chair of the 2nd Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers (SERIAL), Rennes, France, December 2018.

François Taïani served as PC Co-Chair of the 6th Edition of The International Conference on NETworked sYStems (NETYS), Essaouira, Morocco, May 2018.

9.1.2.2. Member of the Conference Program Committees

Pierre-Louis Roman served on the PC of the 16th IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC), Athens, Greece, August 2018.

Pierre-Louis Roman served on the PC of the 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN), Yichang, China, October 2018.

Pierre-Louis Roman served on the PC of the 8th IEEE International Symposium on Cloud and Service Computing (SC2), Paris, France, November 2018.

Pierre-Louis Roman served on the PC of the 11th IEEE International Conference on Service-Oriented Computing and Applications (SOCA), Paris, France, November 2018.

Davide Frey served of the PC of the Industry Track of the 2018 ACM/IFIP International Middleware Conference, Rennes, France, December 2018.

Davide Frey served on the PC of the 18th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS), 2018.

François Taïani served on the PC of the 19th ACM/IFIP/Usenix International Conference on Middleware (Middleware), Rennes, France, December 2018.

François Taïani served on the PC of the 38th IEEE International Conference on Distributed Computing Systems (ICDCS), in the Dependability Track track, Vienna, Austria, July 2018.

François Taïani served on the PC of the Conférence d'informatique en Parallélisme, Architecture et Système (Compas), Toulouse, France, July 2018.

François Taïani served on the PC of The 48th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), Luxembourg, June 2018.

George Giakkoupis served on the PC of he 22nd International Conference on Principles of Distributed Systems (OPODIS), Hong Kong, December 2018.

9.1.3. Journal

9.1.3.1. Reviewer - Reviewing Activities

Anaïs Durand was a reviewer for Theoretical Computer Science (TCS).

Davide Frey was a reviewer for IEEE Transactions on Parallel and Distributed Systems (TPDS).

François Taïani was a reviewer for Journal of Systems Architecture (JSA, Elsevier).

George Giakkoupis was a reviewer for Journal of the ACM (JACM).

George Giakkoupis was a reviewer for Algorithmica (ALGO).

George Giakkoupis was a reviewer for ACM Transactions on Parallel Computing (TOPC).

George Giakkoupis was a reviewer for Theoretical Computer Science (TCS).

9.1.4. Invited Talks

Pierre-Louis Roman. Shortcutting the Bitcoin verification process for your smartphone. EPFL, Lausanne, Switzerland, 1 November 2018.

Pierre-Louis Roman. Shortcutting the Bitcoin verification process for your smartphone. Master Workshop Layer 1 Solutions, Amsterdam, The Netherlands, 18 November 2018.

Anaïs Durand. Contention-Related Crash Failures. Descartes ANR project meeting, IRIF, Université Paris Diderot, Paris, France, 1 October, 2018.

François Taïani. Possible Future Careers after the PhD. Nantes Plenary Meeting and Training School, BigStorage European Training Network (ETN), Nantes, France, 25-28 June 2018.

François Taïani. Pleiades: Distributed Structural Invariants at Scale. Delyss/Whisper Inria Team Retrait, Saint-Valéry, France, 18-20 June 2018.

François Taïani. Pleiades: Distributed Structural Invariants at Scale. Plenary Meeting of the Dionasys Project, Saint-Jacut, France, 23 May 2018.

François Taïani. Being Prepared in a Sparse World: The Case of KNN Graph Construction. EPFL, Lausanne, Switzerland, January 2018.

George Giakkoupis. Random binary search trees with concurrent insertions. Theory Seminar, Technion, Haifa, Israel, 31 October 2018.

George Giakkoupis. Random binary search trees with concurrent insertions. Randomized Algorithms Seminar, University of Cambridge, UK, 2 August 2018.

George Giakkoupis. Random binary search trees with concurrent insertions. ART-Oberseminar, University of Hamburg, Germany, 19 July 2018.

9.1.5. Leadership within the Scientific Community

François Taïani has been a member of the Steering Committee of IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems (DAIS) since 2013.

François Taïani served on the award committee of the Prix de thèse Gilles Kahn 2018, sponsored by the French Science Academy (Académie des Sciences), and awarded by the French Informatics Society (SiF).

Michel Raynal has been Adjunct Professor at the Polytechnic University of Hong Kong since 2013.

Michel Raynal has been a member of the Executive board of the SIF (Société d'Informatique Française) since 2013.

Michel Raynal is European representative at the IEEE Technical Committee on distributed computing.

9.1.6. Research Administration

Davide Frey is Correspondant Scientifique Europe at the DPEI for Inria Rennes

Davide Frey is an associate member of the COST-GTRI of Inria.

François Taïani has been a member of the Scientific Orientation Committee (COS) of IRISA (UMR 6074) since 2013.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence: Olivier Ruas, Spe-Info2, 26h, L1, Université de Rennes 1, France

Licence: Olivier Ruas, Programmation Web, 6h, L2, Université de Rennes 1, France

Master: Davide Frey, Programming Technologies for the Cloud, 28h, M2, Univ. Rennes I, France.

Master: Davide Frey, Scalable Distributed Systems, 10h, M1, EIT/ICT Labs Master School, Univ. Rennes I, France.

Master: Davide Frey, Big-Data Storage and Processing Infrastructures, 10h, M2-SIF, Univ. Rennes I, France.

Master: Davide Frey, Cloud Computing, 6h, M2-MIAGE, Univ. Rennes I, France.

Master: Davide Frey, Distributed Systems, 12h, ENSAI Rennes, France.

Master: George Giakkoupis, Systèmes Répartis, 9h (Spring), ENSAI Rennes, France.

Master: George Giakkoupis, Systèmes Répartis, 9h (Fall), ENSAI Rennes, France.

Engineering School: François Taïani, Synchronization and Parallel Programming, 62h, 2nd year of Engineering School (M1), ESIR / Univ. Rennes I, France.

Engineering School: François Taïani, Distributed Systems, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.

Engineering School: François Taïani, Parallel Algorithms for Big Data, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.

Engineering School: François Taïani, Introduction to Operating Systems, 24h, 1st year of Engineering School (L3), ESIR / Univ. Rennes I, France.

9.2.2. Supervision

PhD: Pierre-Louis Roman, Exploring Heterogeneity in Loosely Consistent Decentralized Data Replication [12], University of Rennes 1, 18 December 2018, François Taïani and Davide Frey.

PhD: Simon Bouget, Towards a Holistic Construction of Opportunistic Large-Scale Distributed Systems [11], University of Rennes 1, 20 September 2018, François Taïani and David Bromberg.

PhD: Olivier Ruas, The Many Faces of Approximation in KNN Graph Computation [13], University of Rennes 1, 17 December 2018, François Taïani and Anne-Marie Kermarrec (Mediego).

PhD in progress: Quentin Dufour, BBDA - Browser Based Data Analytics, January 2018, David Bromberg and Davide Frey.

PhD in progress: Amaury Bouchra Pilet, Robust and Lightweight Overlay Management for Decentralized Learning, University of Rennes 1, September 2018, David Bromberg and Davide Frey.

PhD in progress: Loïck Bonniot, Distributed Troubleshooting of Edge-Compute Functions, University of Rennes 1, François Taïani and Christoph Neumann (Technicolor).

PhD in progress: Hayk Saribekyan, Randomized Algorithms for Distributed Information Dissemination, University of Cambridge, UK, George Giakkoupis and Thomas Sauerwald (Cambridge).

9.2.3. Juries

François Taïani was a reviewer for Rhicheek Patra's PhD thesis: Towards Scalable Personalization, EPFL (Switzerland), 16 January 2018

François Taïani was a reviewer for Raluca Halalaï's PhD thesis: On Reducing Latency in Geo-Distributed Systems through State Partitioning and Caching, Univ. Neuchâtel (Switzerland), 14 May 2018.

François Taïani was a reviewer for William Excoffon's PhD thesis: Résilience des systèmes informatiques adaptatifs : Modélisation, analyse et quantification, Institut National Polytechnique de Toulouse, LAAS, 8 June 2018.

François Taïani was a reviewer for Wenjie Zheng's PhD thesis: A Distributed Frank-Wolfe Framework for Trace Norm Minimization via the Bulk Synchronous Parallel Model, Sorbonne Université, 13 June 2018.

George Giakkoupis was a reviewer for Rémi Varloot's PhD thesis: Dynamic Network Formation. École Polytechnique, Paris, 1 June 2018.

9.3. Popularization

Pierre-Louis Roman gave an invited talk at the E-novation Convention organized by The Notaries of Brittany, Rennes, 12 November 2018. Title: Blockchains, Distributed Ledgers, and their Utility for Notaries (talk in French).

10. Bibliography

Major publications by the team in recent years

[1] P. BERENBRINK, G. GIAKKOUPIS, P. KLING. *Tight Bounds for Coalescing-Branching Random Walks on Regular Graphs*, in "SODA 2018 - Proceedings of the 29th ACM-SIAM Symposium on Discrete Algorithms", New Orleans, United States, ACM, January 2018, pp. 1715-1733, https://hal.inria.fr/hal-01635757

[2] S. BOUGET, Y.-D. BROMBERG, A. LUXEY, F. TAÏANI. *Pleiades: Distributed Structural Invariants at Scale*, in "DSN 2018 - IEEE/IFIP International Conference on Dependable Systems and Networks", Luxembourg, Luxembourg, IEEE, June 2018, pp. 542-553 [*DOI :* 10.1109/DSN.2018.00062], https://hal.archives-ouvertes.fr/hal-01803881

[3] Z. BOUZID, M. RAYNAL, P. SUTRA. *Anonymous obstruction-free (n, k)-set agreement with n-k+1 atomic read/write registers*, in "Distributed Computing", April 2018, vol. 31, n^o 2, pp. 99-117, https://hal.inria.fr/hal-01952626

[4] Y. BROMBERG, Q. DUFOUR, D. FREY. *Multisource Rumor Spreading with Network Coding*, in "2019 IEEE Conference on Computer Communications, INFOCOM 2019, Paris, France, April 29-May 2, 2019", 2019, to appear

[5] Y.-D. BROMBERG, A. LUXEY, F. TAÏANI. *CASCADE: Reliable Distributed Session Handoff for Continuous Interaction across Devices*, in "ICDCS 2018 - 38th IEEE International Conference on Distributed Computing Systems", Vienna, Austria, IEEE, July 2018, pp. 244-254 [*DOI :* 10.1109/ICDCS.2018.00033], https://hal.inria.fr/hal-01797548

[6] K. CENSOR-HILLEL, M. GHAFFARI, G. GIAKKOUPIS, B. HAEUPLER, F. KUHN. *Tight Bounds on Vertex Connectivity Under Sampling*, in "ACM Transactions on Algorithms", May 2017, vol. 13, n^o 2, pp. 19:1 - 19:26 [*DOI :* 10.1145/3086465], https://hal.inria.fr/hal-01635743

[7] F. CHIERICHETTI, G. GIAKKOUPIS, S. LATTANZI, A. PANCONESI. *Rumor Spreading and Conductance*, in "Journal of the ACM (JACM)", August 2018, vol. 65, n^o 4, pp. 17:1-17:21 [*DOI :* 10.1145/3173043], https://hal.inria.fr/hal-01942162

[8] M. HERLIHY, S. RAJSBAUM, M. RAYNAL, J. STAINER. *From wait-free to arbitrary concurrent solo executions in colorless distributed computing*, in "Theoretical Computer Science", June 2017, vol. 683, pp. 1 - 21 [*DOI :* 10.1016/J.TCS.2017.04.007], https://hal.inria.fr/hal-01660566

[9] A. LUXEY, Y.-D. BROMBERG, F. M. COSTA, V. LIMA, R. DA ROCHA, F. TAÏANI. *Sprinkler: A probabilistic dissemination protocol to provide fluid user interaction in multi-device ecosystems*, in "PerCom 2018 - IEEE International Conference on Pervasive Computing and Communications", Athens, Greece, IEEE, March 2018, pp. 1-10 [*DOI :* 10.1109/PERCOM.2018.8444577], https://hal.inria.fr/hal-01704172

[10] B. NÉDELEC, J. TANKE, P. MOLLI, A. MOSTEFAOUI, D. FREY. *An Adaptive Peer-Sampling Protocol for Building Networks of Browsers*, in "World Wide Web", 2017, vol. 25, 1678 p. [*DOI :* 10.1007/s11280-017-0478-5], https://hal.inria.fr/hal-01619906

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] S. BOUGET. *Towards a holistic construction of opportunistic large-scale distributed systems*, Université Rennes 1, September 2018, https://tel.archives-ouvertes.fr/tel-01909849

[12] P.-L. ROMAN. *Exploring heterogeneity in loosely consistent decentralized data replication*, Université de Rennes 1, December 2018, https://hal.inria.fr/tel-01964628

[13] O. RUAS. *The many faces of approximation in KNN graph computation*, Université de rennes 1, December 2018, https://hal.inria.fr/tel-01938076

Articles in International Peer-Reviewed Journals

[14] Z. BOUZID, M. RAYNAL, P. SUTRA. *Anonymous obstruction-free (n,k)-set agreement with n-k+1 atomic read/write registers*, in "Distributed Computing", April 2018, vol. 31, n^{o} 2, pp. 99-117 [*DOI :* 10.1007/S00446-017-0301-7], https://hal.archives-ouvertes.fr/hal-01680833

[15] Z. BOUZID, M. RAYNAL, P. SUTRA. *Anonymous obstruction-free (n, k)-set agreement with n-k+1 atomic read/write registers*, in "Distributed Computing", April 2018, vol. 31, n^{o} 2, pp. 99-117, https://hal.inria.fr/hal-01952626

[16] F. CHIERICHETTI, G. GIAKKOUPIS, S. LATTANZI, A. PANCONESI. *Rumor Spreading and Conductance*, in "Journal of the ACM (JACM)", August 2018, vol. 65, n^{o} 4, pp. 17:1-17:21 [*DOI :* 10.1145/3173043], https://hal.inria.fr/hal-01942162

[17] C. DELPORTE-GALLET, H. FAUCONNIER, S. RAJSBAUM, M. RAYNAL. *Implementing Snapshot Objects on Top of Crash-Prone Asynchronous Message-Passing Systems*, in "IEEE Transactions on Parallel and Distributed Systems", September 2018, vol. 29, n^{o} 9, pp. 2033-2045, https://hal.inria.fr/hal-01955906

[18] H. LAKHLEF, A. BOUABDALLAH, M. RAYNAL, J. BOURGEOIS. *Agent-based broadcast protocols for wireless heterogeneous node networks*, in "Computer Communications", January 2018, vol. 115, pp. 51-63 [*DOI :* 10.1016/J.COMCOM.2017.10.020], https://hal.archives-ouvertes.fr/hal-01714086

[19] A. MOSTEFAOUI, M. RAYNAL, M. ROY. *Time-Efficient Read/Write Register in Crash-prone Asynchronous Message-Passing Systems*, in "Computing", May 2018, pp. 1-16 [*DOI :* 10.1007/S00607-018-0615-8], https://hal.laas.fr/hal-01784210

[20] B. ROSTAMI, A. CHASSEIN, M. HOPF, D. FREY, C. BUCHHEIM, F. MALUCELLI, M. GOERIGK. *The quadratic shortest path problem: complexity, approximability, and solution methods*, in "European Journal of Operational Research", July 2018, vol. 268, n^{o} 2, pp. 473 - 485 [*DOI :* 10.1016/J.EJOR.2018.01.054], https://hal.inria.fr/hal-01781605

International Conferences with Proceedings

[21] P. BERENBRINK, G. GIAKKOUPIS, P. KLING. *Tight Bounds for Coalescing-Branching Random Walks on Regular Graphs*, in "SODA 2018 - Proceedings of the 29th ACM-SIAM Symposium on Discrete Algorithms", New Orleans, United States, ACM, January 2018, pp. 1715-1733, https://hal.inria.fr/hal-01635757

[22] A. BOUTET, F. DE MOOR, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, A. RAULT. *Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter!*, in "DSN 2018 - the 48th International Conference on Dependable Systems and Networks", Luxembourg, Luxembourg, IEEE, June 2018, pp. 466-477 [*DOI :* 10.1109/DSN.2018.00055], https://hal.inria.fr/hal-01787060

[23] Y.-D. BROMBERG, Q. DUFOUR, D. FREY. *Multisource Rumor Spreading with Network Coding*, in "INFOCOM 2019 - IEEE International Conference on Computer Communications", Paris, France, IEEE, April 2019, pp. 1-10, https://hal.inria.fr/hal-01946632

[24] Y.-D. BROMBERG, A. LUXEY, F. TAÏANI. *CASCADE: Reliable Distributed Session Handoff for Continuous Interaction across Devices*, in "ICDCS 2018 - 38th IEEE International Conference on Distributed Computing Systems", Vienna, Austria, IEEE, July 2018, pp. 244-254 [*DOI :* 10.1109/ICDCS.2018.00033], https://hal.inria.fr/hal-01797548

[25] G. GIAKKOUPIS, P. WOELFEL. *An Improved Bound for Random Binary Search Trees with Concurrent Insertions*, in "STACS 2018 - 35th Symposium on Theoretical Aspects of Computer Science", Caen, France, February 2018, pp. 1-13 [*DOI :* 10.4230/LIPIcs.STACS.2018.37], https://hal.inria.fr/hal-01942160

[26] C. HARDY, E. LE MERRER, B. SERICOLA. *Gossiping GANs*, in "DIDL 2018 - Second Workshop on Distributed Infrastructures for Deep Learning", Rennes, France, December 2018 [*DOI :* 10.1145/3286490.3286563], https://hal.inria.fr/hal-01923742

[27] A.-M. KERMARREC, O. RUAS, F. TAÏANI. *Nobody cares if you liked Star Wars: KNN graph construction on the cheap*, in "Europar 2018", Turin, Italy, LNCS, Springer, August 2018, vol. 11014, pp. 419-431 [*DOI :* 10.1007/978-3-319-96983-1_30], https://hal.inria.fr/hal-01867230

[28] D. LIMA, H. MIRANDA, F. TAÏANI. *Weighting Past on the Geo-Aware State Deployment Problem*, in "WoWMoM 2018 - 19th IEEE International Symposium on "A World of Wireless, Mobile and Multimedia Networks"", Chania, Greece, IEEE Computer Society, June 2018, pp. 1–3 [*DOI :* 10.1109/WOWMOM.2018.8449808], https://hal.archives-ouvertes.fr/hal-01943583

[29] A. LUXEY, Y.-D. BROMBERG, F. M. COSTA, V. LIMA, R. DA ROCHA, F. TAÏANI. *Sprinkler: A probabilistic dissemination protocol to provide fluid user interaction in multi-device ecosystems*, in "PerCom 2018 - IEEE International Conference on Pervasive Computing and Communications", Athens, Greece, IEEE, March 2018, pp. 1-10 [*DOI :* 10.1109/PERCOM.2018.8444577], https://hal.inria.fr/hal-01704172

Conferences without Proceedings

[30] K. ALTISEN, S. DEVISMES, A. DURAND. *Acyclic Strategy for Silent Self-Stabilization in Spanning Forests*, in "SSS 2018 - 20th International Symposium on Stabilization, Safety, and Security of Distributed Systems", Tokyo, Japan, LNCS, Springer, November 2018, vol. 11201, pp. 186-202 [*DOI :* 10.1007/978-3-030-03232-6_13], https://hal.archives-ouvertes.fr/hal-01938671

[31] S. BOUGET, Y.-D. BROMBERG, A. LUXEY, F. TAÏANI. *Pleiades: Distributed Structural Invariants at Scale*, in "DSN 2018 - IEEE/IFIP International Conference on Dependable Systems and Networks", Luxembourg, Luxembourg, IEEE, June 2018, pp. 542-553 [*DOI :* 10.1109/DSN.2018.00062], https://hal.archives-ouvertes.fr/hal-01803881

[32] S. BOUGET, Y.-D. BROMBERG, H. MERCIER, E. RIVIÈRE, F. TAÏANI. *Mind the Gap: Autonomous Detection of Partitioned MANET Systems using Opportunistic Aggregation*, in "SRDS 2018 - 37th IEEE International Symposium on Reliable Distributed Systems", Salvador, Brazil, IEEE, October 2018, pp. 1-10, https://hal.archives-ouvertes.fr/hal-01900360

[33] A. DURAND, M. RAYNAL, G. TAUBENFELD. *k-Set Agreement and Renaming in the Presence of Contention-Related Crash Failures*, in "SSS 2018 - 20th International Symposium on Stabilization, Safety, and Security of Distributed Systems", Tokyo, Japan, November 2018, pp. 1-12, https://hal.archives-ouvertes.fr/hal-01938660

Scientific Books (or Scientific Book chapters)

[34] M. RAYNAL. *Fault-Tolerant Message-Passing Distributed Systems*, Springer International Publishing, 2018, https://hal.archives-ouvertes.fr/hal-01951551

Research Reports

[35] A. BOUTET, F. DE MOOR, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, A. RAULT. *Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter!*, Inria Rennes - Bretagne Atlantique, April 2018, pp. 1-12, https://hal.inria.fr/hal-01767059

[36] D. FREY, M. X. MAKKES, P.-L. ROMAN, F. TAÏANI, S. VOULGARIS. *Dietcoin: shortcutting the Bitcoin verification process for your smartphone*, Univ Rennes, Inria, CNRS, IRISA, France ; Vrije Universiteit Amsterdam, The Netherlands ; Athens University of Economics and Business, Greece, March 2018, n^{o} RR-9162, pp. 1-17, https://arxiv.org/abs/1803.10494 , https://hal.inria.fr/hal-01743995

[37] R. GUERRAOUI, A.-M. KERMARREC, O. RUAS, F. TAÏANI. *Fingerprinting Big Data: The Case of KNN Graph Construction*, Inria Rennes - Bretagne Atlantique ; Inria - IRISA - PANAMA ; Université de Rennes 1 ; EPFL ; Mediego, October 2018, n^{o} RR-9218, pp. 1-30, https://hal.inria.fr/hal-01904341

Other Publications

[38] C. HARDY, E. LE MERRER, B. SERICOLA. *MD-GAN: Multi-Discriminator Generative Adversarial Networks for Distributed Datasets*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01946665

References in notes

[39] Y. AFEK, E. GAFNI. *Asynchrony from synchrony*, in "ICDCN", 2013, pp. 225–239

[40] A. AHMED, E. AHMED. *A survey on mobile edge computing*, in "2016 10th International Conference on Intelligent Systems and Control (ISCO)", Jan 2016, pp. 1-8 [*DOI :* 10.1109/ISCO.2016.7727082], https://ieeexplore.ieee.org/document/7727082

[41] T. ALLARD, D. FREY, G. GIAKKOUPIS, J. LEPILLER. *Lightweight Privacy-Preserving Averaging for the Internet of Things*, in "M4IOT 2016 - 3rd Workshop on Middleware for Context-Aware Applications in the IoT", Trento, Italy, ACM, December 2016, pp. 19 - 22 [*DOI :* 10.1145/3008631.3008635], https://hal.inria.fr/hal-01421986

[42] E. ANSHELEVICH, D. CHAKRABARTY, A. HATE, C. SWAMY. *Approximability of the Firefighter Problem: Computing Cuts over Time*, in "Algorithmica", 2012, vol. 62, n^{o} 1-2, pp. 520–536

[43] D. BERNSTEIN. *Containers and Cloud: From LXC to Docker to Kubernetes*, in "IEEE Cloud Computing", Sept 2014, vol. 1, n^{o} 3, pp. 81-84 [*DOI :* 10.1109/MCC.2014.51], https://ieeexplore.ieee.org/document/7036275

[44] M. BERTIER, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, V. LEROY. *The Gossple Anonymous Social Network*, in "ACM/IFIP/USENIX 11th International Middleware Conference (MIDDLEWARE)", Bangalore, India, I. GUPTA, C. MASCOLO (editors), Middleware 2010, Springer, November 2010, vol. LNCS-6452, pp. 191-211 [*DOI :* 10.1007/978-3-642-16955-7_10], https://hal.inria.fr/inria-00515693

[45] F. BONOMI. *Connected vehicles, the internet of things, and fog computing. VANET 2011, 2011*, 2011, Keynote speech at VANET

[46] F. BONOMI, R. MILITO, J. ZHU, S. ADDEPALLI. *Fog Computing and Its Role in the Internet of Things*, in "1 s t MCC Workshop on Mobile Cloud Computing", 2012, http://doi.acm.org/10.1145/2342509.2342513

[47] A. BOUTET, D. FREY, R. GUERRAOUI, A. JÉGOU, A.-M. KERMARREC. *Privacy-Preserving Distributed Collaborative Filtering*, in "Computing", August 2016, vol. 98, n^{o} 8, https://hal.inria.fr/hal-01251314

[48] A. BOUTET, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, R. PATRA. *HyRec: Leveraging Browsers for Scalable Recommenders*, in "Middleware 2014", Bordeaux, France, December 2014 [*DOI :* 10.1145/2663165.2663315], https://hal.inria.fr/hal-01080016

[49] A. BOUTET, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, A. RAULT, F. TAÏANI, J. WANG. *Hide & Share: Landmark-based Similarity for Private KNN Computation*, in "45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)", Rio de Janeiro, Brazil, June 2015, pp. 263-274 [*DOI :* 10.1109/DSN.2015.60], https://hal.archives-ouvertes.fr/hal-01171492

[50] A. BOUTET, D. FREY, A. JÉGOU, A.-M. KERMARREC, H. RIBEIRO. *FreeRec: an Anonymous and Distributed Personalization Architecture*, in "Computing", December 2013, https://hal.inria.fr/hal-00909127

[51] B. COHEN. *Incentives Build Robustness in BitTorrent*, 2003, http://citeseer.ist.psu.edu/cohen03incentives.html

[52] C. DELPORTE-GALLET, H. FAUCONNIER, R. GUERRAOUI, A. TIELMANN. *The disagreement power of an adversary*, in "Distributed Computing", 2011, vol. 24, n^{o} 3-4, pp. 137–147

[53] A. J. DEMERS, D. H. GREENE, C. HAUSER, W. IRISH, J. LARSON, S. SHENKER, H. E. STURGIS, D. C. SWINEHART, D. B. TERRY. *Epidemic Algorithms for Replicated Database Maintenance*, in "PODC", 1987, pp. 1–12

[54] D. FREY, R. GUERRAOUI, A.-M. KERMARREC, M. MONOD, B. KOLDEHOFE, M. MOGENSEN, V. QUÉMA. *Heterogeneous Gossip*, in "Middleware 2009", Urbana-Champaign, IL, United States, December 2009, https://hal.inria.fr/inria-00436125

[55] W. M. GOLAB, V. HADZILACOS, D. HENDLER, P. WOELFEL. *RMR-efficient implementations of comparison primitives using read and write operations*, in "Distributed Computing", 2012, vol. 25, n^{o} 2, pp. 109-162

[56] R. GUERRAOUI, K. HUGUENIN, A.-M. KERMARREC, M. MONOD, S. PRUSTY. *LiFTinG: Lightweight Freerider-Tracking Protocol in Gossip*, in "11th ACM/IFIP/USENIX International Middleware Conference

(MIDDLEWARE)", Bangalore, India, November 2010 [*DOI :* 10.1007/978-3-642-16955-7_16], https://hal.inria.fr/inria-00505268

[57] R. A. HOLLEY, T. M. LIGGETT. *Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model*, in "The Annals of Probability", 1975, vol. 3, n° 4, pp. 643-663

[58] D. IMBS, M. RAYNAL. *A liveness condition for concurrent objects: x-wait-freedom*, in "Concurrency and Computation: Practice and experience", 2011, vol. 23, n° 17, pp. 2154–2166

[59] F. JUNQUEIRA, K. MARZULLO. *A framework for the design of dependent-failure algorithms*, in "Concurrency and Computation: Practice and Experience", 2007, vol. 19, n° 17, pp. 2255–2269

[60] D. KEMPE, J. M. KLEINBERG, É. TARDOS. *Maximizing the spread of influence through a social network*, in "KDD", 2003, pp. 137–146

[61] D. KEMPE, J. M. KLEINBERG, É. TARDOS. *Influential Nodes in a Diffusion Model for Social Networks*, in "ICALP", 2005, pp. 1127–1138

[62] D. KEMPE, J. M. KLEINBERG, É. TARDOS. *Maximizing the Spread of Influence through a Social Network*, in "Theory of Computing", 2015, vol. 11, pp. 105–147

[63] P. KUZNETSOV. *Understanding non-uniform failure models*, in "Bulletin of the EATCS", 2012, n° 106, pp. 53–77

[64] E. LIEBERMAN, C. HAUERT, M. A. NOWAK. *Evolutionary dynamics on graphs*, in "Nature", 2005, vol. 433, n° 7023, pp. 312–316

[65] M. RAYNAL, J. STAINER. *Synchrony weakened by message adversaries vs asynchrony restricted by failure detectors*, in "PODC", Montréal, Canada, Proceedings of the 2013 ACM symposium on Principles of distributed computing, ACM, July 2013, pp. 166-175 [*DOI :* 10.1145/2484239.2484249], https://hal.inria.fr/hal-00920734

[66] N. SANTORO, P. WIDMAYER. *Time is not a healer*, in "Annual Symposium on Theoretical Aspects of Computer Science", Springer, 1989, pp. 304–313

[67] A. VERMA, L. PEDROSA, M. KORUPOLU, D. OPPENHEIMER, E. TUNE, J. WILKES. *Large-scale cluster management at Google with Borg*, in "Tenth European Conference on Computer Systems (Eurosys 2015)", ACM, 2015, 18 p.

[68] L. ZHANG, F. ZHOU, A. MISLOVE, R. SUNDARAM. *Maygh: Building a CDN from Client Web Browsers*, in "8th ACM European Conference on Computer Systems", New York, NY, USA, EuroSys '13, ACM, 2013, pp. 281–294, http://doi.acm.org/10.1145/2465351.2465379