Activity Report 2019

Team WIDE

The World is Distributed: Exploring the Tension between Scale and Coordination

Joint team with Inria Rennes – Bretagne Atlantique

D1 – Large Scale Systems

1 Team, visitors, external collaborators

Preview team list: will be correctly formatted by IRABOT

Francois Taiani, Enseignant [Team leader, Univ de Rennes I, Professor][Habilite]
Davide Frey, Chercheur [Inria, Researcher][Habilite]
George Giakkoupis, Chercheur [Inria, Researcher]
Anne-Marie Kermarrec, Chercheur [Inria, Senior Researcher, from Aug 2019][Habilite]
Erwan Le Merrer, Chercheur [Inria, Advanced Research Position][Habilite]
David Bromberg, Enseignant [Univ de Rennes I, Professor][Habilite]
Michel Raynal, Enseignant [Univ de Rennes I, Emeritus][Habilite]
Noureddine Haouari, PostDoc [Univ de Rennes I, Post-Doctoral Fellow, from Mar 2019]
Alex Auvolat Bernstein, PhD [Univ de Rennes I, PhD Student, from Oct 2019]
Loick Bonniot, PhD [Technicolor, PhD Student, granted by CIFRE]
Amaury Bouchra Pilet, PhD [Univ de Rennes I, PhD Student]
Florestan De Moor, PhD [Ecole normale supérieure de Rennes, PhD Student, from Sep 2019]
Quentin Dufour, PhD [Inria, PhD Student]
Louison Gitzinger, PhD [Univ de Rennes I, PhD Student]
Adrien Luxey, PhD [Univ de Rennes I, PhD Student]
Alex Auvolat Bernstein, Technique [Univ de Rennes I, Engineer, from Mar 2019 until Sep 2019]
Alex Auvolat Bernstein, Stagiaire [Ecole Normale Supérieure Paris, until Feb 2019]
Florestan De Moor, Stagiaire [Ecole normale supérieure de Rennes, from Feb 2019 until Jun 2019]
Josselin Giet, Stagiaire [Ecole Normale Supérieure Paris, from Sep 2019]
Charaf Sfaoua, Stagiaire [IRISA, from Jun 2019 until Jul 2019]
Hasnaa Dyani, Visiteur [ENSIAS, Rabat, stagiaire, from Apr 2019 until Jun 2019]
Arsany Guirguis, Visiteur [EPFL, Lausanne, PhD student, from Jul 2019 until Sep 2019]
Lionel Kaboret, Visiteur [Joseph Ki-Zerbo University, Ouagadougou, PhD Student, from Sep 2019 until Oct 2019]
Mohamed Lechiakh, Visiteur [ENSIAS, Rabat, stagiaire, from Mar 2019 until May 2019]
Roberto Rodrigues Filho, Visiteur [Lancaster University, from Jul 2019 until Sep 2019]
Chaimaa Tarzi, Visiteur [ENSIAS, Rabat, stagiaire, from Apr 2019 until Jun 2019]

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

On-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 Scientific Foundations

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,

[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.

- 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.

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 (Challenrges 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 [BFG+10,FGK+09] 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) [BFG+14]. In the short term, we aim to explore two axes in this direction: browser-based communication, and micro services.

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

[BFG+10] 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)*, I. Gupta, C. Mascolo (editors), *Middleware 2010*, *LNCS-6452*, Springer, p. 191–211, Bangalore, India, November 2010, https://hal.inria.fr/inria-00515693.

[FGK+09] D. Frey, R. Guerraoui, A.-M. Kermarrec, M. Monod, K. Boris, M. Martin, V. Quéma, "Heterogeneous Gossip", *in: Middleware 2009*, Urbana-Champaign, IL, United States, December 2009, https://hal.inria.fr/inria-00436125.

[BFG+14] A. Boutet, D. Frey, R. Guerraoui, A.-M. Kermarrec, R. Patra, "HyRec: Leveraging Browsers for Scalable Recommenders", *in: Middleware 2014*, Bordeaux, France, December 2014, https://hal.inria.fr/hal-01080016.

traditional cloud model. In 2011 Cisco [Bon11] 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 [AA16].

A number of efforts in this directions focus on specific hardware, e.g. fog nodes that are responsible for connected IoT devices [BMZA12]. 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 [ZZMS13], 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?

[Bon11] F. Bonomi, "Connected vehicles, the internet of things, and fog computing. VANET 2011, 2011", Keynote speech at VANET, 2011.

[AA16] A. Ahmed, E. Ahmed, "A survey on mobile edge computing", *in: 2016 10th International Conference on Intelligent Systems and Control (ISCO)*, p. 1–8, Jan 2016.

[BMZA12] F. Bonomi, R. Milito, J. Zhu, S. Addepalli, "Fog Computing and Its Role in the Internet of Things", *in: 1st MCC Workshop on Mobile Cloud Computing*, 2012, http://doi.acm.org/10.1145/2342509.2342513.

[ZZMS13] L. Zhang, F. Zhou, A. Mislove, R. Sundaram, "Maygh: Building a CDN from Client Web Browsers", *in: 8th ACM European Conference on Computer Systems, EuroSys '13*, ACM, p. 281–294, New York, NY, USA, 2013, http://doi.acm.org/10.1145/2465351.2465379.

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 [VPK+15], Kubernetes [Ber14]) 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

[VPK+15] 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, p. 18, 2015.

[Ber14] D. Bernstein, "Containers and Cloud: From LXC to Docker to Kubernetes", *IEEE Cloud Computing 1*, 3, Sept 2014, p. 81–84.

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).

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 [BFG+16,BFG+15,BFJ+13] as well as the results of our work on large-scale hybrid architectures in Objective 1.

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 [Coh03,GHK+10]. 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.

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 Ob-

[BFG+16] A. Boutet, D. Frey, R. Guerraoui, A. Jégou, A.-M. Kermarrec, "Privacy-Preserving Distributed Collaborative Filtering", *Computing 98*, 8, August 2016, https://hal.inria.fr/hal-01251314.

[BFG+15] 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)*, p. 263–274, Rio de Janeiro, Brazil, June 2015, https://hal.archives-ouvertes.fr/hal-01171492.

[BFJ+13] A. Boutet, D. Frey, A. Jégou, A.-M. Kermarrec, H. Ribeiro, "FreeRec: an Anonymous and Distributed Personalization Architecture", *Computing*, December 2013, https://hal.inria.fr/hal-00909127.

[Coh03] B. Cohen, "Incentives Build Robustness in BitTorrent", 2003, http://citeseer.ist.psu.edu/cohen03incentives.html.

[GHK+10] 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, https://hal.inria.fr/inria-00505268.

jective 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 [AFGL16].

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.

[AFGL16] 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*, ACM, p. 19 – 22, Trento, Italy, December 2016, `https://hal.inria.fr/hal-01421986`.

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.

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 [KKT03,KKT05,KKT15] that have been studied in this context so far. This includes the *randomized rumor spreading (RS)* model for information dissemination [DGH+87], *biased* versions of the *voter model* [HL75] modelling influence, and the (graph-based) *Moran processes* [LHN05] 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 [KKT03], and will also explore new approaches.

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 [ACHS12]. 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

[KKT03] D. Kempe, J. M. Kleinberg, É. Tardos, "Maximizing the spread of influence through a social network", *in: KDD*, p. 137–146, 2003.

[KKT05] D. Kempe, J. M. Kleinberg, É. Tardos, "Influential Nodes in a Diffusion Model for Social Networks", *in: ICALP*, p. 1127–1138, 2005.

[KKT15] D. Kempe, J. M. Kleinberg, É. Tardos, "Maximizing the Spread of Influence through a Social Network", *Theory of Computing 11*, 2015, p. 105–147.

[DGH+87] 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*, p. 1–12, 1987.

[HL75] R. A. Holley, T. M. Liggett, "Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model", *The Annals of Probability 3*, 4, 1975, p. 643–663.

[LHN05] E. Lieberman, C. Hauert, M. A. Nowak, "Evolutionary dynamics on graphs", *Nature 433*, 7023, 2005, p. 312–316.

[ACHS12] E. Anshelevich, D. Chakrabarty, A. Hate, C. Swamy, "Approximability of the Firefighter Problem: Computing Cuts over Time", *Algorithmica 62*, 1-2, 2012, p. 520–536.

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*.

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 [GHHW12]. 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.

[GHHW12] W. M. Golab, V. Hadzilacos, D. Hendler, P. Woelfel, "RMR-efficient implementations of comparison primitives using read and write operations", *Distributed Computing 25*, 2, 2012, p. 109–162.

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.

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 [SW89]) and *process adversaries* (investigated in several papers, e.g. [RS13,DGFGT11,IR11,JM07,Kuz12]). 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

[SW89] N. Santoro, P. Widmayer, "Time is not a healer", *in : Annual Symposium on Theoretical Aspects of Computer Science*, Springer, p. 304–313, 1989.

[RS13] M. Raynal, J. Stainer, "Synchrony weakened by message adversaries vs asynchrony restricted by failure detectors", *in : PODC, Proceedings of the 2013 ACM symposium on Principles of distributed computing*, ACM, p. 166–175, Montréal, Canada, July 2013, `https://hal.inria.fr/hal-00920734`.

[DGFGT11] C. Delporte-Gallet, H. Fauconnier, R. Guerraoui, A. Tielmann, "The disagreement power of an adversary", *Distributed Computing 24*, 3-4, 2011, p. 137–147.

[IR11] D. Imbs, M. Raynal, "A liveness condition for concurrent objects: x-wait-freedom", *Concurrency and Computation: Practice and experience 23*, 17, 2011, p. 2154–2166.

[JM07] F. Junqueira, K. Marzullo, "A framework for the design of dependent-failure algorithms", *Concurrency and Computation: Practice and Experience 19*, 17, 2007, p. 2255–2269.

[Kuz12] P. Kuznetsov, "Understanding non-uniform failure models", *Bulletin of the EATCS*, 106, 2012, p. 53–77.

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 [RS13,AG13].

4 Highlights of the Year

4.1 Awards

- Florestan De Moor is the recipient of the "Prix National Jeunes André Blanc-Lapierre 2019" from the SEE society (Société de l'électricité, de l'électronique et des technologies de l'information et de la communication), for his work during his master thesis [25].
- During the SRDS 2019 conference which took place in Lyon, France, from October 1st to 4th, Michel Raynal received an Outstanding Career Award for his contributions to distributed systems and algorithms.

5 New Results

5.1 Recommender Systems

5.1.1 A Biclustering Approach to Recommender Systems

Participants: Florestan De Moor, Davide Frey.

Recommendation systems are a core component of many e-commerce industries and online services since they ease the discovery of relevant products. Because catalogs are huge, it is impossible for an individual to manually search for an item of interest, hence the need for some automatic filtering process. Many approaches exist, from content-based ones to collaborative filtering that include neighborhood and model-based techniques. Despite these intensive research activities, numerous challenges remain to be addressed, particularly under real-time settings or regarding privacy concerns, which motivates further work in this area. We focus on techniques that rely on biclustering, which consists in simultaneously building clusters over the two dimensions of a data matrix. Although it was little considered by the recommendation system community, it is a well-known technique in other domains such as genomics. In work [25] we present the different biclustering-based approaches that were explored. We then are the first to perform an extensive experimental evaluation to compare these approaches with

[RS13] M. Raynal, J. Stainer, "Synchrony weakened by message adversaries vs asynchrony restricted by failure detectors", *in: PODC, Proceedings of the 2013 ACM symposium on Principles of distributed computing*, ACM, p. 166–175, Montréal, Canada, July 2013, https://hal.inria.fr/hal-00920734.

[AG13] Y. Afek, E. Gafni, "Asynchrony from synchrony", *in: ICDCN*, p. 225–239, 2013.

one another, but also with the current state-of-the-art techniques from the recommender field. Existing evaluations are often restrained to a few algorithms and consider only a limited set of metrics. We then expose a few ideas to improve existing approaches and address the current challenges in the design of highly efficient recommendation algorithms, along with some preliminary results.

This work was done in collaboration with Antonio Mucherino (University of Rennes 1).

5.1.2 Unified and Scalable Incremental Recommenders with Consumed Item Packs

Participants: Erwan Le Merrer.

Recommenders personalize the web content by typically using collaborative filtering to relate users (or items) based on explicit feedback, e.g., ratings. The difficulty of collecting this feedback has recently motivated to consider implicit feedback (e.g., item consumption along with the corresponding time). In this work [17], we introduce the notion of consumed itempack (CIP) which enables to link users (or items) based on their implicit analogous consumption behavior. Our proposal is generic, and we show that it captures three novel implicit recommenders: a user-based (CIP-U), an item-based (CIP-I),and a word embedding-based (DEEP-CIP), as well as a state-of-art technique using implicit feedback (FISM). We show that our recommenders handle incremental updates incorporating freshly consumed items. We demonstrate that all three recommenders provide a recommendation quality that is competitive with state-of-the-art ones, including one incorporating both explicit and implicit feedback

This work was done in collaboration with Rachid Guerraoui (EPFL), Rhicheek Patra (Oracle) and Jean-Ronan Vigouroux (Technicolor).

5.2 Systems for the Support of Privacy

5.2.1 Robust Privacy-Preserving Gossip Averaging

Participants: Amaury Bouchra-Pilet, Davide Frey, François Taïani.

This contribution aims to address the privacy risks inherent in decentralized systems by considering the emblematic problem of privacy-preserving decentralized averaging. In particular, we propose a novel gossip protocol that exchanges noise for several rounds before starting to exchange actual data. This makes it hard for an honest but curious attacker to know whether a user is transmitting noise or actual data. Our protocol and analysis do not assume a lock-step execution, and demonstrate improved resilience to colluding attackers. In a paper, publishing this work at SSS 2019 [12], we prove the correctness of this protocol as well as several privacy results. Finally, we provide simulation results about the efficiency of our averaging protocol.

5.2.2 A Collaborative Strategy for Mitigating Tracking through Browser Fingerprinting.

Participants: David Bromberg, Davide Frey, Alejandro Gomez-Boix.

Browser fingerprinting is a technique that collects information about the browser configuration and the environment in which it is running. This information is so diverse that it can partially or totally identify users online. Over time, several countermeasures have emerged to mitigate tracking through browser fingerprinting. However, these measures do not offer full coverage in terms of privacy protection, as some of them may introduce inconsistencies or unusual behaviors, making these users stand out from the rest.

In this work, we address these limitations by proposing a novel approach that minimizes both the identifiability of users and the required changes to browser configuration. To this end, we exploit clustering algorithms to identify the devices that are prone to share the same or similar fingerprints and to provide them with a new non-unique fingerprint. We then use this fingerprint to automatically assemble and run web browsers through virtualization within a docker container. Thus all the devices in the same cluster will end up running a web browser with an indistinguishable and consistent fingerprint.

We carried out this work in collaboration with Benoit Baudry from KTH Sweden and published our results at the 2019 Moving-Target Defense Workshop [18].

5.3 Distributed Algorithms

5.3.1 One for All and All for One:Scalable Consensus in a Hybrid Communication Model

Participants: Michel Raynal.

This work [22] addresses consensus in an asynchronous model where the processes are partitioned into clusters. Inside each cluster, processes can communicate through a shared memory, which favors efficiency. Moreover, any pair of processes can also communicate through a message-passing communication system, which favors scalability. In such a "hybrid communication" context, the work presents two simple binary consensus algorithms (one based on local coins,the other one based on a common coin). These algorithms are straightforward extensions of existing message-passing randomized round-based consensus algorithms. At each round, the processes of each cluster first agree on the same value (using an underlying shared memory consensus algorithm), and then use a message-passing algorithm to converge on the same decided value. The algorithms are such that, if all except one processes of a cluster crash, the surviving process acts as if all the processes of its cluster were alive (hence the motto "one for all and all for one"). As a consequence, the hybrid communication model allows us to obtain simple, efficient, and scalable fault-tolerant consensus algorithms. As an important side effect, according to the size of each cluster, consensus can be obtained even if a majority of processes crash.

This work was done in collaboration with Jiannong Cao (Polytechnic University, Hong Kong).

5.3.2 Optimal Memory-Anonymous Symmetric Deadlock-Free Mutual Exclusion

Participants: Michel Raynal.

The notion of an anonymous shared memory (recently introduced in PODC 2017) considers that processes use different names for the same memory location. Hence, there is permanent disagreement on the location names among processes. In this context, the PODC paper presented -among other results- a symmetric deadlock-free mutual exclusion (mutex) algorithm for two processes and a necessary condition on the size m of the anonymous memory for the existence of a symmetric deadlock-free mutex algorithm in an n-process system. This work [9] answers several open problems related to symmetric deadlock-free mutual exclusion in an n-process system where the processes communicate through m registers. It first presents two algorithms. The first considers that the registers are anonymous read/write atomic registers and works for any m greater than 1 and belonging to the set M(n). It thus shows that this condition on m is both necessary and sufficient. The second algorithm considers anonymous read/modify/write atomic registers. It assumes that $m \in M(n)$. These algorithms differ in their design principles and their costs (measured as the number of registers which must contain the identity of a process to allow it to enter the critical section). The work also shows that the condition $m \in M(n)$ is necessary for deadlock-free mutex on top of anonymous read/modify/write atomic registers. It follows that, when m > 1, $m \in M(n)$ is a tight characterization of the size of the anonymous shared memory needed to solve deadlock-free mutex, be the anonymous registers read/write or read/modify/write.

This work was done in collaboration with Zahra Aghazadeh (University of Calgary), Damien Imbs (LIS, Université d'Aix-Marseille,CNRS, Université de Toulon), Gadi Taubenfeld (The Interdisciplinary Center of Herzliya) and Philipp Woelfel (University of Calgary).

5.3.3 Merkle Search Trees

Participants: Alex Auvolat, François Taïani.

Most recent CRDT (Conflict-free Replicated Data Type) techniques rely on a causal broadcast primitive to provide guarantees on the delivery of operation deltas. Such a primitive is unfortunately hard to implement efficiently in large open networks, whose membership is often difficult to track. As an alternative, we argue that pure state-based CRDTs can be efficiently implemented by encoding states as specialized Merkle trees, and that this approach is well suited to open networks where many nodes may join and leave. Indeed, Merkle trees enable efficient remote comparison and reconciliation of data sets, which can be used to implement the CRDT merge operator between two nodes without any prior information. This approach also does not require vector clock information, which would grow linearly with the number of participants.

At the core of our contribution [11] lies a new kind of Merkle tree, called Merkle Search Tree (MST), that implements a balanced search tree while maintaining key ordering. This

latter property makes it particularly efficient in the case of updates on sets of sequential keys, a common occurrence in many applications. We use this new data structure to implement a distributed event store, and show its efficiency in very large systems with low rates of updates. In particular, we show that in some scenarios our approach is able to achieve both a 66% reduction of bandwidth cost over a vector-clock approach, as well as a 34% improvement in consistency level. We finally suggest other uses of our construction for distributed databases in open networks.

5.3.4 Dietcoin: Hardening Bitcoin Transaction Verification Process For Mobile Devices

Participants: Davide Frey, François Taïani.

Distributed ledgers are among the most replicated data repositories in the world. They offer data consistency, immutability, and auditability, based on the assumption that each participating node locally verifies their entire content. Although their content, currently extending up to a few hundred gigabytes, can be accommodated by dedicated commodity hard disks, downloading it, processing it, and storing it in general-purpose desktop and laptop computers can prove largely impractical. Even worse, this becomes a prohibitive restriction for smartphones, mobile devices, and resource-constrained IoT devices.

We thus proposed Dietcoin, a Bitcoin protocol extension that allows nodes to perform secure local verification of Bitcoin transactions with small bandwidth and storage requirements. We carried out an extensive evaluation of the features of Dietcoin that are important for today's cryptocurrency and smart-contract systems, but are missing in the current state-of-the-art. These include (i) allowing resource-constrained devices to verify the correctness of selected blocks locally without having to download the complete ledger; (ii) enabling devices to join a blockchain quickly yet securely, dropping bootstrap time from days down to a matter of seconds; (iii) providing a generic solution that can be applied to other distributed ledgers secured with Proof-of-Work. We showcased our results in a demo at VLDB 2019 [3], and we are currently preparing a full paper submission.

We carried out this work in collaboration with Pierre-Louis Roman, now at University of Lugano (Switzerland), as well as with Mark Makke from Vrije Universiteit, Amsterdam (the Netherlands), and Spyros Voulgaris from Athens University of Economics and Business (Greece).

5.3.5 Byzantine-Tolerant Set-Constrained Delivery Broadcast

Participants: Alex Auvolat, François Taïani, Michel Raynal.

Set-Constrained Delivery Broadcast (SCD-broadcast), recently introduced at ICDCN 2018, is a high-level communication abstraction that captures ordering properties not between individual messages but between sets of messages. More precisely, it allows processes to broadcast messages and deliver sets of messages, under the constraint that if a process delivers a set containing a message m before a set containing a message m', then no other process delivers first a set containing m' and later a set containing m. It has been shown that SCD-broadcast

and read/write registers are computationally equivalent, and an algorithm implementing SCD-broadcast is known in the context of asynchronous message passing systems prone to crash failures.

We introduce a Byzantine-tolerant SCD-broadcast algorithm in [10], which we call BSCD-broadcast. Our proposed algorithm assumes an underlying basic Byzantine-tolerant reliable broadcast abstraction. We first introduce an intermediary communication primitive, Byzantine FIFO broadcast (BFIFO-broadcast), which we then use as a primitive in our final BSCD-broadcast algorithm. Unlike the original SCD-broadcast algorithm that is tolerant to up to $t < n/2$ crashing processes, and unlike the underlying Byzantine reliable broadcast primitive that is tolerant to up to $t < n/3$ Byzantine processes, our BSCD-broadcast algorithm is tolerant to up to $t < n/4$ Byzantine processes. As an illustration of the high abstraction power provided by the BSCD-broadcast primitive, we show that it can be used to implement a Byzantine-tolerant read/write snapshot object in an extremely simple way.

5.3.6 PnyxDB: a Lightweight Leaderless Democratic Byzantine Fault Tolerant Replicated Datastore

Participants: Loïck Bonniot, François Taïani.

Byzantine-Fault-Tolerant (BFT) systems are rapidly emerging as a viable technology for production-grade systems, notably in closed consortia deployments for financial and supply-chain applications. Unfortunately, most algorithms proposed so far to coordinate these systems suffer from substantial scalability issues, mainly due to the requirement of a single leader node. We observed that many application workloads offer little concurrency, and proposed PnyxDB, an eventually-consistent BFT replicated datastore that exhibits both high scalability and low latency. Our approach (proposed in [23]) is based on conditional endorsements, that allow nodes to specify the set of transactions that must *not* be committed for the endorsement to be valid.

Additionally, although most of prior art rely on internal voting or quorum mechanisms, these mechanisms are not exposed to applications as first-class primitives. As a result, individual nodes cannot implement application-defined policies without additional effort, costs, and complexity. This is problematic, as application-level voting capabilities are key to a number of emerging decentralized BFT applications involving independent participants who need to balance conflicting goals and shared interests. In addition to its high scalability, PnyxDB supports application-level voting by design. We provided a comparison against BFTSMaRt and Tendermint, two competitors with different design aims, and demonstrated that our implementation speeds up commit latencies by a factor of 11, remaining below 5 seconds in a worldwide geodistributed deployment of 180 nodes.

PnyxDB's source code is freely available[2]. This work has also been done in collaboration with Christoph Neumann at InterDigital.

[2] `https://github.com/technicolor-research/pnyxdb`

5.3.7 Vertex Coloring with Communication Constraints in Synchronous Broadcast Networks

Participants: Hicham Lakhlef, Michel Raynal, François Taïani.

In this work [5], we consider distributed vertex-coloring in broadcast/receive networks suffering from conflicts and collisions. (A collision occurs when, during the same round, messages are sent to the same process by too many neighbors; a conflict occurs when a process and one of its neighbors broadcast during the same round.) More specifically, our work focuses on multi-channel networks, in which a process may either broadcast a message to its neighbors or receive a message from at most γ of them. The work first provides a new upper bound on the corresponding graph coloring problem (known as frugal coloring) in general graphs, proposes an exact bound for the problem in trees, and presents a deterministic, parallel, color-optimal, collision- and conflict-free distributed coloring algorithm for trees, and proves its correctness.

5.3.8 Efficient Randomized Test-and-Set Implementations

Participants: George Giakkoupis.

In [4], we study randomized test-and-set (TAS) implementations from registers in the asynchronous shared memory model with n processes. We introduce the problem of *group election*, a natural variant of leader election, and propose a framework for the implementation of TAS objects from group election objects. We then present two group election algorithms, each yielding an efficient TAS implementation. The first implementation has expected maxstep complexity $O(\log^* k)$ in the location-oblivious adversary model, and the second has expected maxstep complexity $O(\log\log k)$ against any read/write-oblivious adversary, where $k \leq n$ is the contention. These algorithms improve the previous upper bound by Alistarh and Aspnes (2011) of $O(\log\log n)$ expected maxstep complexity in the oblivious adversary model.

We also propose a modification to a TAS algorithm by Alistarh, Attiya, Gilbert, Giurgiu, and Guerraoui (2010) for the strong adaptive adversary, which improves its space complexity from super-linear to linear, while maintaining its $O(\log n)$ expected maxstep complexity. We then describe how this algorithm can be combined with any randomized TAS algorithm that has expected maxstep complexity $T(n)$ in a weaker adversary model, so that the resulting algorithm has $O(\log n)$ expected maxstep complexity against any strong adaptive adversary and $O(T(n))$ in the weaker adversary model.

Finally, we prove that for any randomized 2-process TAS algorithm, there exists a schedule determined by an oblivious adversary such that with probability at least $1/4^t$ one of the processes needs at least t steps to finish its TAS operation. This complements a lower bound by Attiya and Censor-Hillel (2010) on a similar problem for $n \geq 3$ processes.

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

5.4 Machine Learning and Security

5.4.1 Adversarial Frontier Stitching for Remote Neural Network Watermarking

Participants: Erwan Le Merrer.

The state-of-the-art performance of deep learning models comes at a high cost for companies and institutions, due to the tedious data collection and the heavy processing requirements. Recently, Nagai et al. proposed to watermark convolutional neural networks for image classification, by embedding information into their weights. While this is a clear progress toward model protection, this technique solely allows for extracting the watermark from a network that one accesses locally and entirely. Instead, we aim at allowing the extraction of the watermark from a neural network (or any other machine learning model) that is operated remotely, and available through a service API. To this end, we propose in this work [6] to mark the model's action itself, tweaking slightly its decision frontiers so that a set of specific queries convey the desired information. In this work, we formally introduce the problem and propose a novel zero-bit watermarking algorithm that makes use of adversarial model examples. While limiting the loss of performance of the protected model, this algorithm allows subsequent extraction of the watermark using only few queries. We experimented the approach on three neural networks designed for image classification, in the context of the MNIST digit recognition task.

This work was done in collaboration with Gilles Trédan (LAAS/CRNS) and Patrick Pérez (Valéo AI).

5.4.2 TamperNN: Efficient Tampering Detection of Deployed Neural Nets

Participants: Erwan Le Merrer.

Neural networks are powering the deployment of embedded devices and Internet of Things. Applications range from personal assistants to critical ones such as self-driving cars. It has been shown recently that models obtained from neural nets can be trojaned ; an attacker can then trigger an arbitrary model behavior facing crafted inputs. This has a critical impact on the security and reliability of those deployed devices. In this work [21], we introduce novel algorithms to detect the tampering with deployed models, classifiers in particular. In the remote interaction setup we consider, the proposed strategy is to identify markers of the model input space that are likely to change class if the model is attacked, allowing a user to detect a possible tampering. This setup makes our proposal compatible with a wide range of scenarios, such as embedded models, or models exposed through prediction APIs. We experiment those tampering detection algorithms on the canonical MNIST dataset, over three different types of neural nets, and facing five different attacks (trojaning, quantization, fine-tuning, compression and watermarking). We then validate over five large models (VGG16, VGG19, ResNet, MobileNet, DenseNet) with a state of the art dataset (VGGFace2), and report results demonstrating the possibility of an efficient detection of model tampering.

This work was done in collaboration with Gilles Trédan (LAAS/CRNS).

5.4.3 MD-GAN: Multi-Discriminator Generative Adversarial Networks for Distributed Datasets

Participants: Erwan Le Merrer.

A recent technical breakthrough in the domain of machine learning is the discovery and the multiple applications of Generative Adversarial Networks (GANs). Those generative models are computationally demanding, as a GAN is composed of two deep neural networks, and because it trains on large datasets. A GAN is generally trained on a single server. In this work, we address the problem of distributing GANsso that they are able to train over datasets that are spread on multiple workers. In this work [19] MD-GAN is exposed as the first solution for this problem: we propose a novel learning procedure for GANs so that they fit this distributed setup. We then compare the performance of MD-GAN to an adapted version of Federated Learning to GANs, using the MNIST and CIFAR10 datasets.MD-GAN exhibits a reduction by a factor of two of the learning complexity on each worker node, while providing better performances than federated learning on both datasets. We finally discuss the practical implications of distributing GANs.

This work was done in collaboration with Bruno Sericola (Inria) and Corentin Hardy (Technicolor).

5.5 Network and Graph Algorithms

5.5.1 Multisource Rumor Spreading with Network Coding

Participants: David Bromberg, Quentin Dufour, Davide Frey.

The last decade has witnessed a rising interest in Gossip protocols in distributed systems. In particular, as soon as there is a need to disseminate events, they become a key functional building block due to their scalability, robustness and fault tolerance under high churn. However, Gossip protocols are known to be bandwidth intensive. A huge amount of algorithms has been studied to limit the number of exchanged messages using different combinations of push/pull approaches. In this work we revisited the state of the art by applying Random Linear Network Coding to further increase performance. In particular, the originality of our approach consists in combining sparse-vector encoding to send our network-coding coefficients and Lamport timestamps to split messages in generations in order to provide efficient gossiping. Our results demonstrate that we are able to drastically reduce bandwidth overhead and dissemination delay compared to the state of the art. We published our results at INFOCOM 2019 [13].

5.5.2 DiagNet: towards a generic, Internet-scale root cause analysis solution

Participants: Loïck Bonniot, François Taïani.

Internet content providers and network operators allocate significant resources to diagnose and troubleshoot problems encountered by end-users, such as service quality of experience degradations. Because the Internet is decentralized, the cause of such problems might lie anywhere between an end-user's device and the service datacenters. Further, the set of possible problems and causes cannot be known in advance, making it impossible to train a classifier with all combinations of faults, causes and locations. We explored how machine learning can

be used for Internet-scale root cause analysis using measurements taken from end-user devices: our solution, DiagNet, is able to build generic models that (i) do not make any assumption on the underlying network topology, (ii) do not require to define the full set of possible causes during training, and (iii) can be quickly adapted to diagnose new services.

DiagNet adapts recent image analysis tactics for system and network metrics, collected from a large and dynamic set of landmark servers. In details, it applies non-overlapping convolutions and global pooling to extract generic information about the analyzed network. This genericness allows to build a general model, that can later be generalized to any Internet service with minimal effort. DiagNet leverages backpropagation attention mechanisms to extend the possible root causes to the set of available metrics, making the model fully extensible. We evaluated DiagNet on geodistributed mockup web services and automated users running in 6 AWS regions, and demonstrated promising root cause analysis capabilities. While this initial work is being reviewed, we are deploying DiagNet for real web services and users to evaluate its performance in a more realistic setup.

Christoph Neumann (InterDigital) actively participated in this work.

5.5.3 Application-aware adaptive partitioning for graph processing systems

Participants: Erwan Le Merrer.

Modern online applications value real-time queries over fresh data models. This is the case for graph-based applications, such as social networking or recommender systems,running on front-end servers in production. A core problem in graph processing systems is the efficient partitioning of the input graph over multiple workers. Recent advances over Bulk Synchronous Parallel processing systems (BSP) enabled computations over partitions on those workers, independently of global synchronization supersteps. A good objective partitioning makes the understanding of the load balancing and communication trade-off mandatory for performance improvement. This work [20] addresses this trade-off through the proposal of an optimization problem, that is to be solved continuously to avoid performance degradation over time. Our simulations show that the design of the software module we propose yields significant performance improvements over the BSP processing model.

This work was done in collaboration with Gilles Trédan (LAAS/CRNS).

5.5.4 How to Spread a Rumor: Call Your Neighbors or Take a Walk?

Participants: George Giakkoupis.

In [15], we study the problem of randomized information dissemination in networks. We compare the now standard push-pull protocol, with agent-based alternatives where information is disseminated by a collection of agents performing independent random walks. In the visit-exchange protocol, both nodes and agents store information, and each time an agent visits a node, the two exchange all the information they have. In the meet-exchange protocol, only the agents store information, and exchange their information with each agent they meet.

We consider the broadcast time of a single piece of information in an n-node graph for the above three protocols, assuming a linear number of agents that start from the stationary distri-

bution. We observe that there are graphs on which the agent-based protocols are significantly faster than push-pull, and graphs where the converse is true. We attribute the good performance of agent-based algorithms to their inherently fair bandwidth utilization, and conclude that, in certain settings, agent-based information dissemination, separately or in combination with push-pull, can significantly improve the broadcast time.

The graphs considered above are highly non-regular. Our main technical result is that on any regular graph of at least logarithmic degree, push-pull and visit-exchange have the same asymptotic broadcast time. The proof uses a novel coupling argument which relates the random choices of vertices in push-pull with the random walks in visit-exchange. Further, we show that the broadcast time of meet-exchange is asymptotically at least as large as the other two's on all regular graphs, and strictly larger on some regular graphs.

As far as we know, this is the first systematic and thorough comparison of the running times of these very natural information dissemination protocols.

This work was done in collaboration with Frederik Mallmann-Trenn (MIT) and Hayk Saribekyan (University of Cambridge, UK).

6 Contracts and Grants with Industry

6.1 Bilateral Contracts with Industry

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

Participants: Loïck Bonniot, 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 Partnerships and Cooperations

7.1 Regional Initiatives

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

Participants: 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.

7.2 National Initiatives

7.2.1 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.

7.2.2 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.

7.2.3 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.

7.2.4 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.

7.3 International Initiatives

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 is a step in this inescapable direction.

7.4 International Research Visitors

- Roberto Rodrigues Filho (Lancaster University, UK), July–September 2019.
- Mohamed Lechiakh, (ENSIAS, Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes, Rabat, Morocco), March–May 2019.
- Hasnaa Dyani, (ENSIAS, Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes, Rabat, Morocco), April–June 2019.
- Chaimaa Tarzi, (ENSIAS, Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes, Rabat, Morocco), April–June 2019.
- Arsany Guirguis, (EPFL, Lausanne, Switzerland), July–September 2019.
- Marcus Kaboret, (Laboratoire de Mathématiques et Informatique, Joseph Ki-Zerbo University, Ouagadougou, Burkina Faso), September–October 2019.
- Hayk Saribekyan (University of Cambridge, UK), 2-–12 April 2019.
- Giorgi Nadiradze (IST Austria), 20–24 May 2019.
- Emanuele Natale (CNRS, Sophia-Antipolis), 13–17 March 2019.

7.5 Visits to International Teams

- Adrien Luxey visited Paulo Ferreira, University of Oslo, Norway, from the 1st May to the 30th of June 2019.

8 Dissemination

8.1 Promoting Scientific Activities

Member of the Organizing Committees

- François Taïani and Erwan Le Merrer served as Co-Organizer of the 9th INRIA/Interdigtal Workshop On Systems (WOS9), Rennes, France, December 2019.
- François Taïani served as Posters and Demos Co-Chair of the 38th IEEE International Symposium on Reliable Distributed Systems (SRDS 2019), Lyon, France, October 2019.
- Erwan Le Merrer served as Co-Organizer of the "Atelier sur les algorithmes en boite-noire", Lyon, France, December 2019.

8.2 Scientific Events Selection

8.2.1 Chair of Conference Program Committees

- François Taïani served as PC Co-Chair of the 3^{rd} Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers (SERIAL'19), collocated with Middleware 2019, Davis, CA, USA, December 2019.
- François Taïani served as Co-Chair of the Shadow PC of the European Conference on Computer Systems (EuroSys 2019), Dresden, Germany, March 2019.
- Erwan Le Merrer served as Co-Chair of the "Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications" (Algotel 2019), France, June 2019.

8.2.2 Member of the Conference Program Committees

- François Taïani served on the PC of the Conférence d'informatique en Parallélisme, Architecture et Système (COMPAS2019), Anglet, France, June 2019.
- François Taïani served on the PC of the 18th Workshop on Adaptive and Reflexive Middleware (ARM'19@MW), co-located with Middleware 2019, Davis, CA, USA, December 2019.
- François Taïani served on the PC of the 49th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2019), Portaland, OR, USA, June 2019.
- François Taïani served on the PC of the International Conference on Blockchain Economics, Security and Protocols (Tokenomics 2019), Paris, France, May 2019.
- Davide Frey served on the PC of the 5th Workshop on Planetary-Scale Distributed Systems (W-PSDS), co-located with SRDS 2019, Lyon, France, October 2019.
- Davide Frey served on the PC of the 13th ACM International Conference on Distributed and Event-Based Systems (DEBS), Darmstadt, Germany, June 2019.
- Davide Frey served on the PC of the 19th International Conference on Distributed Applications and Interoperable Systems (DAIS), Copenhagen, Denmark, June 2019.
- George Giakkoupis served on the PC of the 31st ACM Symposium on Parallelism in Algorithms and Architectures (SPAA), Phoenix, AZ, USA, June 22–24, 2019.
- George Giakkoupis served on the PC of the 33rd International Symposium on Distributed Computing (DISC), Budapest, Hungary, Oct. 14–18, 2019.

8.3 Journal

Reviewer - Reviewing Activities

- François Taïani was a reviewer for IEEE Transactions on Knowledge and Data Engineering (TKDE).

- Davide Frey was a reviewer for the International Journal of Computer and Telecommunications Networking (COMNET).
- Davide Frey was a reviewer for Peer-to-Peer Networking and Applications (PPNA).
- Davide Frey was a reviewer for the ARIMA Journal.
- George Giakkoupis was a reviewer for Journal of the ACM (JACM).
- George Giakkoupis was a reviewer for Journal of Parallel and Distributed Computing (JPDC).
- George Giakkoupis was a reviewer for ACM Transactions on Parallel Computing (TOPC).
- George Giakkoupis was a reviewer for Distributed Computing (DIST).

8.4 Collaborative Projects

- Davide Frey was a reviewer for the Agence Nationale de Recherche (ANR).

8.5 Invited Talks

- François Taïani. Pleiades: Distributed Structural Invariants at Scale. 4th GDR RSD and ASF Winter School on Distributed Systems and Networks 2019, Pleynet, Sept Laux, France, 6 February 2019.
- François Taïani. Pleiades: Distributed Structural Invariants at Scale. Lancaster University, Lancaster, UK, 23 May 2019.
- Erwan Le Merrer. Tweaking neural models: watermarking and tamperproofing them. EPFL-Inria workshop, EPFL, Lausanne, 30 January 2019.
- François Taïani. How to Be a Terrible Higher Education Teacher in 7 Easy Steps. IRISA Annual D1 Departmental Seminar, IRISA, Rennes, 29 November 2019.

8.6 Leadership within the Scientific Community

- 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.
- François Taïani has been a member of the Gilles Kahn PhD Award in Computer Science, from Société Informatique de France (SIF) since 2018.

8.7 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 is a Career Advice Person, (*Référent conseil-parcours professionel chercheurs*) for IRISA/Inria Rennes Bretagne Atlantique since 2019.

8.8 Teaching - Supervision - Juries

8.8.1 Teaching

- Master: Florestan De Moor, Programmation Dirigée par la Syntaxe, 22h, M1-SIF, Université de Rennes 1, 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, Introduction to Operating Systems, 24h, 1st year of Engineering School (L3), ESIR / Univ. Rennes I, France.
- Bacherlor: François Taïani, Distributed Algorithms, 12h, L3 Parcours SI, ENS Rennes, France.
- Master: Davide Frey, Programming Technologies for the Cloud, 28h, M2, Univ. Rennes I, France.
- Master: Davide Frey, Scalable Distributed Systems, 10 hours, M1, EIT/ICT Labs Master School, Univ. Rennes I, France.
- Master: Davide Frey, Big-Data Storage and Processing Infrastructures, 10 hours, M2-SIF, Univ. Rennes I, France.
- Master: Davide Frey, Cloud Computing, 6 hours, M2-MIAGE, Univ. Rennes I, France.
- Master: Davide Frey, Distributed Systems, 12 hours, ENSAI, France.
- Master: Davide Frey, Apprentice Tutoring, 8 ETD hours, M2 Alternance Univ. Rennes I, France.
- Master: Quentin Dufour and Davide Frey, 12 ETD hours, M1 ISTIC, Univ. Rennes I, France.

- Master / PhD: Davide Frey, Distributed Computing and Blockchain, 15 hours, UM6P, Morocco.

- Master: Erwan Le Merrer, Projet , 36 ETD hours, M1 ISTIC, Univ. Rennes I, France.

- Master: Erwan Le Merrer, Network Science, 12 hours, M2 ESIR, Univ. Rennes I, France.

- Master: Quentin Dufour and Louison Gitzinger, Cloud, 36 hours, M2 ESIR, Univ. Rennes I, France.

- Master: Quentin Dufour, TLC, 18 hours, M2 ISTIC, Univ. Rennes I, France

- Master: George Giakkoupis, Systèmes Répartis, 9 hours, ENSAI Rennes, France.

8.8.2 Supervision

- PhD: Adrien Luxey, "E-squads: A novel paradigm to build privacy-preserving ubiquitous applications" [2], University of Rennes 1, 29th November 2019, David Bromberg.

- 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 (Interdigital, CIFRE).

- PhD in progress: Alejandro Gomez Boix, Distributed counter-measure against browser fingerprinting, Inria, Davide Frey and David Bromberg.

- PhD in progress: Alex Auvolat, Towards probabilistic decentralized systematic design for large-scale privacy-preserving collaborative systems, University of Rennes 1, François Taïani and David Bromberg.

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

- HDR: Davide Frey, Epidemic Protocols: From Large Scale to Big Data, University of Rennes 1, 11 Juin 2019 [1].

8.8.3 Juries

- François Taïani was an examiner for Cédric Maigrot's PhD thesis: Détection de fausses informations dans les réseaux sociaux, Université de Rennes 1 (France), 1^{st} April 2019.
- François Taïani was a reviewer for João Paulo de Araujo's PhD thesis: A Communication-Efficient Causal Broadcast Publish/Subscribe System, Sorbonne Université, LIP6 (France), 5^{th} April 2019.
- François Taïani was an external examiner (reviewer) for Assylbek Sagitzhanuly Jumagaliyev's PhD thesis: A Modeling Language for Multi-tenant Data Architecture Evolution in Cloud Applications, Lancaster University (UK), 23^{rd} May 2019.
- François Taïani was an examiner for Gilles Tredan's HDR: Capturing Binary Graphs, Université de Toulouse 3 (France), 21^{st} June 2019.
- François Taïani was an examiner for Nathanaël Cheriere's PhD thesis: Towards Malleable Distributed Storage Systems: from Models to Practice, ENS Rennes, 5^{th} November 2019.
- François Taïani was an examiner for The Anh Pham's PhD thesis: Efficient state-space exploration for asynchronous distributed programs, ENS Rennes, 6^{th} December 2019.
- François Taïani was an examiner for Oscar Luis Vera Perez's PhD thesis: Dynamic program analysis for suggesting test improvements to developers, University of Rennes 1, 17^{th} December 2019.
- George Giakkoupis was an external reviewer for Suman Sourav's PhD thesis: Latency, Conductance, and the Role of Connectivity in Graph Problems, National University of Singapore, 11 November 2019.

8.9 Popularization

- Algorithm Watch wrote an article about a recent work involving Erwan Le Merrer, concernant remote transparency. The article is called "Explainable AI doesn't work for online services – now there's proof". Avaiable at `https://algorithmwatch.org/en/story/explainable-ai-doesnt-work-for-online-services-now-theres-proof/`.

9 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*, ACM, p. 1715–1733, New Orleans, United States, January 2018, `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*, IEEE, p. 542–553, Luxembourg, Luxembourg, June 2018, 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", *Distributed Computing 31*, 2, April 2018, p. 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*, IEEE, p. 244–254, Vienna, Austria, July 2018, 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", *ACM Transactions on Algorithms 13*, 2, May 2017, p. 19:1 – 19:26, https://hal.inria.fr/hal-01635743.

[7] F. Chierichetti, G. Giakkoupis, S. Lattanzi, A. Panconesi, "Rumor Spreading and Conductance", *Journal of the ACM (JACM) 65*, 4, August 2018, p. 17:1–17:21, 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", *Theoretical Computer Science 683*, June 2017, p. 1 – 21, https://hal.inria.fr/hal-01660566.

[9] H. Lakhlef, M. Raynal, F. Taïani, "Vertex Coloring with Communication Constraints in Synchronous Broadcast Networks", *IEEE Transactions on Parallel and Distributed Systems 30*, 7, July 2019, p. 1672–1686, https://hal.inria.fr/hal-02376726.

[10] 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*, IEEE, p. 1–10, Athens, Greece, March 2018, https://hal.inria.fr/hal-01704172.

[11] B. Nédelec, J. Tanke, P. Molli, A. Mostefaoui, D. Frey, "An Adaptive Peer-Sampling Protocol for Building Networks of Browsers", *World Wide Web 25*, 2017, p. 1678, https://hal.inria.fr/hal-01619906.

Doctoral dissertations and "Habilitation" theses

[1] D. Frey, *Epidemic Protocols: From Large Scale to Big Data*, Habilitation à diriger des recherches, Université De Rennes 1, June 2019, https://hal.inria.fr/tel-02375909.

[2] A. Luxey, *E-squads: A novel paradigm to build privacy-preserving ubiquitous applications*, Thèse, Université de Rennes, November 2019, https://hal.inria.fr/tel-02389297.

Articles in referred journals and book chapters

[3] D. Frey, M. X. Makkes, P.-L. Roman, F. Taïani, S. Voulgaris, "Dietcoin: Hardening Bitcoin Transaction Verification Process For Mobile Devices", *Proceedings of the VLDB Endowment (PVLDB) 12*, 12, August 2019, p. 1946–1949, https://hal.inria.fr/hal-02315154.

[4] G. Giakkoupis, P. Woelfel, "Efficient Randomized Test-And-Set Implementations", *Distributed Computing*, 2019, p. 565–586, https://hal.inria.fr/hal-02012672.

[5] H. Lakhlef, M. Raynal, F. Taïani, "Vertex Coloring with Communication Constraints in Synchronous Broadcast Networks", *IEEE Transactions on Parallel and Distributed Systems 30*, 7, July 2019, p. 1672–1686, https://hal.inria.fr/hal-02376726.

[6] E. Le Merrer, P. Pérez, G. Trédan, "Adversarial frontier stitching for remote neural network watermarking", *Neural Computing and Applications*, 2019, p. 1–12, https://hal.archives-ouvertes.fr/hal-02264449.

[7] A. Mostefaoui, M. Perrin, M. Raynal, J. Cao, "Crash-Tolerant Causal Broadcast in O(n) Messages", *Information Processing Letters 151*, November 2019, p. 1–9, https://hal.archives-ouvertes.fr/hal-02279523.

[8] A. Mostefaoui, M. Raynal, M. Roy, "Time-Efficient Read/Write Register in Crash-prone Asynchronous Message-Passing Systems", *Computing 101*, 1, January 2019, p. 3–17, https://hal.laas.fr/hal-01784210.

Publications in Conferences and Workshops

[9] Z. Aghazadeh, D. Imbs, M. Raynal, G. Taubenfeld, P. Woelfel, "Optimal Memory-Anonymous Symmetric Deadlock-Free Mutual Exclusion", *in : PODC*, Toronto, Canada, July 2019, https://hal.archives-ouvertes.fr/hal-02394246.

[10] A. Auvolat, M. Raynal, F. Taïani, "Byzantine-Tolerant Set-Constrained Delivery Broadcast", *in : OPODIS 2019 - International Conference on Principles of Distributed Systems*, ACM, p. 1–23, Neuchâtel, Switzerland, December 2019, https://hal.inria.fr/hal-02376673.

[11] A. Auvolat, F. Taïani, "Merkle Search Trees: Efficient State-Based CRDTs in Open Networks", *in : SRDS 2019 - 38th IEEE International Symposium on Reliable Distributed Systems*, IEEE, p. 1–10, Lyon, France, October 2019, https://hal.inria.fr/hal-02303490.

[12] A. Bouchra Pilet, D. Frey, F. Taïani, "Robust Privacy-Preserving Gossip Averaging", *in : SSS 2019 - 21st International Symposium on Stabilization, Safety, and Security of Distributed Systems*, Springer, p. 38–52, Pisa, Italy, November 2019, https://hal.archives-ouvertes.fr/hal-02373353.

[13] Y.-D. Bromberg, Q. Dufour, D. Frey, "Multisource Rumor Spreading with Network Coding", *in : INFOCOM 2019 - IEEE International Conference on Computer Communications*, IEEE, p. 1–10, Paris, France, April 2019, https://hal.inria.fr/hal-01946632.

[14] A. Durand, M. Raynal, G. Taubenfeld, "Pannes de processus liées à la contention", *in : ALGOTEL 2019 - 21èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications*, p. 1–4, Saint Laurent de la Cabrerisse, France, June 2019, https://hal.archives-ouvertes.fr/hal-02118917.

[15] G. Giakkoupis, F. Mallmann-Trenn, H. Saribekyan, "How to Spread a Rumor: Call Your Neighbors or Take a Walk?", *in : PODC 2019 - ACM Symposium on Principles of Distributed Computing*, ACM Press, p. 24–33, Toronto ON, Canada, July 2019, https://hal.inria.fr/hal-02388328.

[16] R. Guerraoui, A.-M. Kermarrec, O. Ruas, F. Taïani, "Fingerprinting Big Data: The Case of KNN Graph Construction", *in : ICDE 2019 - 35th IEEE International Conference on Data Engineering*, IEEE, p. 1738–1741, Macao, China, April 2019, https://hal.inria.fr/hal-02357950.

[17] R. Guerraoui, E. Le Merrer, R. Patra, J.-R. Vigouroux, "Unified and Scalable Incremental Recommenders with Consumed Item Packs", *in : EURO-PAR 2019 - European Conference on Parallel Processing*, Springer, p. 227–240, Gottingen, Germany, August 2019, https://hal.archives-ouvertes.fr/hal-02153388.

[18] A. Gómez-Boix, D. Frey, Y.-D. Bromberg, B. Baudry, "A Collaborative Strategy for mitigating Tracking through Browser Fingerprinting", *in : MTD 2019 - 6th ACM Workshop on Moving Target Defense*, p. 1–12, London, United Kingdom, November 2019, https://hal.inria.fr/hal-02282591.

[19] C. Hardy, E. Le Merrer, B. Sericola, "MD-GAN: Multi-Discriminator Generative Adversarial Networks for Distributed Datasets", *in : IPDPS 2019 - 33rd IEEE International Parallel and Distributed Processing Syposium*, IEEE, p. 1–12, Rio de Janeiro, Brazil, May 2019, https://hal.inria.fr/hal-01946665.

[20] E. Le Merrer, G. Trédan, "Application-aware adaptive partitioning for graph processing systems", *in : MASCOTS 2019 - 27th IEEE International Symposium on the Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, IEEE, p. 235–240, Rennes, France, October 2019, https://hal.archives-ouvertes.fr/hal-02193594.

[21] E. Le Merrer, G. Trédan, "TamperNN: Efficient Tampering Detection of Deployed Neural Nets", *in : ISSRE 2019 - IEEE 30th International Symposium on Software Reliability Engineering*, p. 1–11, Berlin, Germany, October 2019, https://hal.archives-ouvertes.fr/hal-02268136.

[22] M. Raynal, J. Cao, "One for All and All for One: Scalable Consensus in a Hybrid Communication Model", *in : ICDCS*, Dallas, United States, July 2019, https://hal.archives-ouvertes.fr/hal-02394259.

Miscellaneous

[23] L. Bonniot, C. Neumann, F. Taïani, "PnyxDB: a Lightweight Leaderless Democratic Byzantine Fault Tolerant Replicated Datastore", https://arxiv.org/abs/1911.03291 - working paper or preprint, November 2019, https://hal.archives-ouvertes.fr/hal-02355778.

[24] A. Bouchra Pilet, D. Frey, F. Taïani, "Simple, Efficient and Convenient Decentralized Multi-Task Learning for Neural Networks", working paper or preprint, November 2019, https://hal.archives-ouvertes.fr/hal-02373338.

[25] F. De Moor, *A Biclustering Approach to Recommender Systems*, Mémoire, University of Rennes 1, June 2019, https://hal.inria.fr/hal-02369708.