Activity Report 2018

Team DELYS

DistributEd aLgorithms and sYStems

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Paris

THEME
Distributed Systems and middleware

Table of contents

Team DELYS

Creation of the Team: 2018 January 01, updated into Project-Team: 2019 January 01

Keywords:

Computer Science and Digital Science:

A1.1.1. - Multicore, Manycore
A1.1.9. - Fault tolerant systems
A1.2.5. - Internet of things
A1.3.2. - Mobile distributed systems
A1.3.3. - Blockchain
A1.3.4. - Peer to peer
A1.3.5. - Cloud
A1.3.6. - Fog, Edge
A1.5.2. - Communicating systems
A2.6. - Infrastructure software
A2.6.1. - Operating systems
A2.6.2. - Middleware
A2.6.3. - Virtual machines
A2.6.4. - Ressource management
A3.1.8. - Big data (production, storage, transfer)
A7.1.1. - Distributed algorithms

Other Research Topics and Application Domains:

B6.4. - Internet of things

1. Team, Visitors, External Collaborators

Research Scientists

Mesaac Makpangou [Inria, Researcher, HDR]
Marc Shapiro [Inria, Senior Researcher, HDR]

Faculty Members

Luciana Bezerra Arantes [Sorbonne Univ, Associate Professor]
Philippe Darche [Univ René Descartes, Associate Professor]
Swan Dubois [Sorbonne Univ, Associate Professor]
Jonathan Lejeune [Sorbonne Univ, Associate Professor]
Franck Petit [Sorbonne Univ, Professor, HDR]
Pierre Sens [Team leader, Sorbonne Univ, Professor, HDR]
Julien Sopena [Sorbonne Univ, Associate Professor]

Post-Doctoral Fellows

Laurent Feuilloley [Sorbonne Univ]
Paolo Viotti [Sorbonne Univ, until Feb. 2018]

PhD Students

Sébastien Bouchard [Inria]
Marjorie Bournat [Sorbonne Univ]
Michael Damien Carver [Magency, Sorbonne Univ]

João Paulo de Araujo [Sorbonne Univ]
Arnaud Favier [Inria, from Oct 2018]
Guillaume Fraysse [Orange Labs]
Lyes Hamidouche [Magency, until Apr 2018]
Saalik Hatia [Sorbonne Univ, since Oct 2018]
Denis Jeanneau [Sorbonne Univ]
Francis Laniel [Sorbonne Univ]
Sreeja Nair [Sorbonne Univ, since April 2018]
Jonathan Sid-Otmane [Orange Labs, Sorbonne Univ]
Alejandro Tomsic [Inria, until Apr 2018]
Ilyas Toumlilt [Sorbonne Univ]
Dimitrios Vasilas [Scality, Sorbonne Univ]
Gauthier Voron [Sorbonne Univ, until Apr 2018]

Administrative Assistant
Nelly Maloisel [Inria]

External Collaborator
Sébastien Monnet [Univ Savoie Mont-Blanc]

2. Overall Objectives

2.1. Overall Objectives

The research of the Delys team addresses the theory and practice of distributed systems, including multicore computers, clusters, networks, peer-to-peer systems, cloud and fog computing systems, and other communicating entities such as swarms of robots. It addresses the challenges of correctly communicating, sharing information, and computing in such large-scale, highly dynamic computer systems. This includes addressing the core problems of communication, consensus and fault detection, scalability, replication and consistency of shared data, information sharing in collaborative groups, dynamic content distribution, and multi- and many-core concurrent algorithms.

Delys is a joint research team between LIP6 (UPMC/CNRS) and Inria Paris.

3. Research Program

3.1. Research rationale

DELYS addresses both theoretical and practical issues of *Computer Systems*, leveraging our dual expertise in theoretical and experimental research. Our approach is a “virtuous cycle,” triggered by issues with real systems, of algorithm design which we prove correct and evaluate theoretically, and then implement and test experimentally feeding back to theory. The major challenges addressed by DELYS are the sharing of information and guaranteeing correct execution of highly-dynamic computer systems. Our research covers a large spectrum of distributed computer systems: multicore computers, mobile networks, cloud computing systems, and dynamic communicating entities. This holistic approach enables handling related problems at different levels. Among such problems we can highlight consensus, fault detection, scalability, search of information, resource allocation, replication and consistency of shared data, dynamic content distribution, and concurrent and parallel algorithms.

Two main evolutions in the Computer Systems area strongly influence our research project:

(1) Modern computer systems are **increasingly distributed**, **dynamic** and composed of multiple devices **geographically spread over heterogeneous platforms**, spanning multiple management domains. Years of research in the field are now coming to fruition, and are being used by millions of users of web systems, peer-to-peer systems, gaming and social applications, cloud computing, and now fog computing. These new uses bring new challenges, such as *adaptation to dynamically-changing conditions*, where knowledge of the system state can only be partial and incomplete.

(2) **Heterogeneous architectures and virtualisation are everywhere**. The parallelism offered by distributed clusters and *multicore* architectures is opening highly parallel computing to new application areas. To be successful, however, many issues need to be addressed. Challenges include obtaining a consistent view of shared resources, such as memory, and optimally distributing computations among heterogeneous architectures. These issues arise at a more fine-grained level than before, leading to the need for different solutions down to OS level itself.

The scientific challenges of the distributed computing systems are subject to many important features which include scalability, fault tolerance, dynamics, emergent behaviour, heterogeneity, and virtualisation at many levels. Algorithms designed for traditional distributed systems, such as resource allocation, data storage and placement, and concurrent access to shared data, need to be redefined or revisited in order to work properly under the constraints of these new environments. Sometimes, classical "*static*" problems, (*e.g.*, Election Leader, Spanning Tree Construction, ...) even need to be redefined to consider the unstable nature of the distributed system. In particular, DELYS will focus on three key challenges:

Rethinking distributed algorithms. From a theoretical point of view the key question is how to adapt the fundamental building blocks to new architectures. More specifically, how to rethink the classical algorithms to take into account the dynamics of advanced modern systems. Since a recent past, there have been several papers that propose models for dynamic systems: there is practically a different model for each setting and currently there is no unification of models. Furthermore, models often suffer of lack of realism. One of the key challenge is to identify which assumptions make sense in new distributed systems. DELYS's objectives are then (1) to identify under which realistic assumptions a given fundamental problem such as mutual exclusion, consensus or leader election can be solved and (2) to design efficient algorithms under these assumptions.

Resource management in heterogeneous systems. The key question is how to manage resources on large and heterogeneous configurations. Managing resources in such systems requires fully decentralized solutions, and to rethink the way various platforms can collaborate and interoperate with each other. In this context, data management is a key component. The fundamental issue we address in ow to efficiently and reliably share information in highly distributed environments.

Adaptation of runtimes. One of the main challenge of the OS community is how to adapt runtime supports to new architectures. With the increasingly widespread use of multicore architectures and virtualised environments, internal runtime protocols need to be revisited. Especially, memory management is crucial in OS and virtualisation technologies have highly impact on it. On one hand, the isolation property of virtualisation has severe side effects on the efficiency of memory allocation since it needs to be constantly balanced between hosted OSs. On the other hand, by hiding the physical machine to OSs, virtualisation prevents them to efficiently place their data in memory on different cores. Our research will thus focus on providing solutions to efficiently share memory between OSs without jeopardizing isolation properties.

4. Highlights of the Year

4.1. Highlights of the Year

In 2018, the DELYS team published papers at major conferences in Systems, Distributed Systems, Theoretical Computer Science, Verification, and AI:

- Scheduling under Uncertainty: A Query-based Approach. L. Arantes, E. Bampis, A. Kononov, M. Letsios, G. Lucarelli, P. Sens. IJCAI, [19].
- Byzantine Gathering in Polynomial Time. S. Bouchard, Y. Dieudonné, A. Lamani. ICALP [22].
- The Battle of the Schedulers: FreeBSD ULE vs. Linux CFS. J. Bouron, S. Chevalley, B. Lepers, W. Zwaenepoel, R. Gouicem, J. Lawall, G. Muller, J. Sopena. ATC [24].
- Distributed transactional reads: the strong, the quick, the fresh & the impossible. A. Z. Tomsic, M. Bravo, M. Shapiro. Middleware [31].
- Co-design and verification of an available file system. M. Najafzadeh, M. Shapiro, P. Eugster. VMCAI [28].

5. New Results

5.1. Distributed Algorithms for Dynamic Networks and Fault Tolerance

Participants: Luciana Bezerra Arantes [correspondent], Sébastien Bouchard, Marjorie Bournat, João Paulo de Araujo, Swan Dubois, Laurent Feuilloley, Denis Jeanneau, Jonathan Lejeune, Franck Petit [correspondent], Pierre Sens, Julien Sopena.

Nowadays, distributed systems are more and more heterogeneous and versatile. Computing units can join, leave or move inside a global infrastructure. These features require the implementation of *dynamic* systems, that is to say they can cope autonomously with changes in their structure in terms of physical facilities and software. It therefore becomes necessary to define, develop, and validate distributed algorithms able to managed such dynamic and large scale systems, for instance mobile *ad hoc* networks, (mobile) sensor networks, P2P systems, Cloud environments, robot networks, to quote only a few.

The fact that computing units may leave, join, or move may result of an intentional behavior or not. In the latter case, the system may be subject to disruptions due to component faults that can be permanent, transient, exogenous, evil-minded, etc. It is therefore crucial to come up with solutions tolerating some types of faults.

In 2018, we obtained the following results.

5.1.1. Scheduling in uncertain environments

In [19], we consider scheduling with faults/errors and we introduce a new non-probabilistic model with explorable (query-able) uncertainty. Each unit-time error is characterized by an uncertainty area during which the error will occur, and it is possible to learn the exact slot at which it will appear by issuing a query operation of unit cost. We study two problems: (i) the error-query scheduling problem, whose aim is to reveal enough error-free slots with the minimum number of queries, and (ii) the lexicographic error-query scheduling problem where we seek the earliest error-free slots with the minimum number of queries. We consider both the off-line and the on-line versions of the above problems. In the former, the whole instance and its characteristics are known in advance and we give a polynomial-time algorithm for the error-query scheduling problem. In the latter, the adversary has the power to decide, in an on-line way, the time-slot of appearance for each error. We propose then both lower bounds and algorithms whose competitive ratios asymptotically match these lower bounds.

5.1.2. Failure detectors in dynamic systems

The failure detector abstraction was introduced as a way to circumvent the impossibility of solving consensus in asynchronous systems prone to crash failures. A failure detector is a local oracle that provides processes in the system with unreliable information on process failures. But a failure detector that is sufficient to solve a given problem in a static system is not necessarily sufficient to solve the same problem in a dynamic system. In [37], we adapt an existing failure detector for mutual exclusion and prove that it is the weakest failure detector to solve mutual exclusion in dynamic systems, which means that it is weaker than any other failure detector capable of solving mutual exclusion.

We also propose in [15] a new failure detector, called the Impact failure detector (FD), that expresses the confidence with regard to the system as a whole. Similarly to a reputation approach, it is possible to indicate the relative importance of each process of the system, while a threshold offers a degree of flexibility for failures and false suspicions. Performance evaluation results, based on real PlanetLab traces, confirm the degree of flexible of the failure detector.

5.1.3. Causal information dissemination

A causal broadcast ensures that messages are delivered to all nodes (processes) preserving causal relation of the messages. In [33], we propose a new causal broadcast protocol for distributed systems whose nodes are logically organized in a virtual hypercube-like topology called VCube. Messages are broadcast by dynamically building spanning trees rooted in the message's source node. By using multiple trees, the contention bottleneck problem of a single root spanning tree approach is avoided. Furthermore , different trees can intersect at some node. Hence, by taking advantage of both the out-of-order reception of causally related messages at a node and these paths intersections, a node can delay to one or more of its children in the tree. Experimental evaluation conducted on top of PeerSim simulator confirms the communication effectiveness of our causal broadcast protocol in terms of latency and message traffic reduction

5.1.4. Graceful Degradation

Gracefully degrading algorithms was introduced by Biely *et al.*i. Such algorithms offer the desirable properties to circumvent impossibility results in dynamic systems by adapting themselves to the dynamics. Indeed, such a algorithms solve a given problem under some dynamics and, moreover, guarantees that a weaker (but related) problem is solved under a higher dynamics under which the original problem is impossible to solve. The underlying intuition is to solve the problem whenever possible but to provide some kind of quality of service if the dynamics become (unpredictably) higher.

In [36], we apply for the first time this approach to robot networks. We focus on the fundamental problem of gathering a squad of autonomous robots on an unknown location of a dynamic ring. In this goal, we introduce a set of weaker variants of this problem. Motivated by a set of impossibility results related to the dynamics of the ring, we propose a gracefully degrading gathering algorithm.

5.1.5. Unreliable Hints

In [23], we address the question of a mobile agent deterministically searching for a target in the Euclidean plane. We assume that the mobile agent is equipped with a compass and a measure of length has to find an inert treasure in the Euclidean plane. Both the agent and the treasure are modeled as points. In the beginning, the agent is at a distance at most $D > 0$ from the treasure, but knows neither the distance nor any bound on it. Finding the treasure means getting at distance at most 1 from it. The agent makes a series of moves. Each of them consists in moving straight in a chosen direction at a chosen distance. In the beginning and after each move the agent gets a hint consisting of a positive angle smaller than 2π whose vertex is at the current position of the agent and within which the treasure is contained. We investigate the problem of how these hints permit the agent to lower the cost of finding the treasure, using a deterministic algorithm, where the cost is the worst-case total length of the agent's trajectory. It is well known that without any hint the optimal (worst case) cost is $\Theta(D^2)$. We show that if all angles given as hints are at most π, then the cost can be lowered to $O(D)$, which is optimal. If all angles are at most β, where $\beta < 2\pi$ is a constant unknown to the agent, then the cost is at most $O(D^2 - \epsilon)$, for some $\epsilon > 0$. For both these positive results we present deterministic algorithms achieving the above costs. Finally, if angles given as hints can be arbitrary, smaller than 2π, then we show that cost $\Theta(D^2)$ cannot be beaten.

5.1.6. Gathering of Mobile Agents

Gathering a group of mobile agents is a fundamental task in the field of distributed and mobile systems. It consists of bringing agents that initially start from different positions to meet all together in finite time. In the case when there are only two agents, the gathering problem is often referred to as the rendezvous problem.

In [14] and [22], we consider these tasks from a deterministic point of view in networks modeled as undirected and anonymous graphs. An adversary chooses the initial nodes of the agents (the number of agents may be larger than the number of nodes) and assigns a different positive integer (called label) to each of them. Initially, each agent knows its label as well as some global knowledge shared by all the agents. The agents can communicate with each other only when located at the same node.

This task has been considered in the literature under two alternative scenarios: weak and strong. Under the weak scenario, agents may meet either at a node or inside an edge. Under the strong scenario, they have to meet at a node, and they do not even notice meetings inside an edge. Gathering and rendezvous algorithms under the strong scenario are known for synchronous agents. For asynchronous agents, gathering and rendezvous under the strong scenario are impossible even in the two-node graph, and hence only algorithms under the weak scenario were constructed.

In [14] we show that rendezvous under the strong scenario is possible for agents with asynchrony restricted in the following way: agents have the same measure of time but the adversary can impose, for each agent and each edge, the speed of traversing this edge by this agent. The speeds may be different for different edges and different agents but all traversals of a given edge by a given agent have to be at the same imposed speed. We construct a deterministic rendezvous algorithm for such agents, working in time polynomial in the size of the graph, in the length of the smaller label, and in the largest edge traversal time.

Gathering mobile agents can be made drastically more difficult to achieve when some agents are subject to faults, especially the Byzantine ones that are known as being the worst faults to handle. Byzantine means that the agent is subject to unpredictable and arbitrary faults. For instance, such an agent may choose to never stop or to never move. In [22] we study the task of Byzantine gathering among synchronous agents under the strong scenario: despite the presence of f Byzantine agents, all the other (correct) agents have to meet at the same node. In this respect, assuming that the agents are in a *strong team* i.e., a team in which the number of correct agents is at least some prescribed value that is quadratic in f, we show an algorithm that solves Byzantine gathering with all strong teams in all graphs of size at most n, for any integers n and f, in a time polynomial in n and the length $|l_{min}|$ of the binary representation of the smallest label of a good agent. The algorithm works using a global knowledge of size $\mathcal{O}(\log \log \log n)$, which we prove to be of optimal order of magnitude in our context to reach a time complexity that is polynomial in n and $|l_{min}|$.

5.1.7. Self-Stabilizing Minimum Diameter Spanning Tree

In [13], we present a self-stabilizing algorithm for the minimum diameter spanning tree construction problem in the state model. Our protocol has the following attractive features. It is the first algorithm for this problem that operates under the *unfair and distributed* adversary (or *daemon*). In other words, no restriction is made on the asynchronous behavior of the system. Second, our algorithm needs only $O(\log n)$ bits of memory per process (where n is the number of processes), that improves the previous result by a factor n. These features are not achieved to the detriment of the convergence time, which stays polynomial.

5.2. Large-scale data distribution

Participants: Saalik Hatia, Mesaac Makpangou, Sébastien Monnet, Sreeja Nair, Jonathan Sid-Otmane, Pierre Sens, Marc Shapiro, Alejandro Tomsic, Ilyas Toumlilt, Dimitrios Vasilas, Paolo Viotti.

5.2.1. Impossibility results for distributed transactional reads

We study the costs and trade-offs of providing transactional consistent reads in a distributed storage system. We identify the following dimensions: read consistency, read delay (latency), and data freshness. We show that there is a three-way trade-off between them, which can be summarised as follows: (i) it is not possible to ensure at the same time order-preserving (e.g., causally-consistent) or atomic reads, Minimal Delay, and maximal freshness; thus, reading data that is the most fresh without delay is possible only in a weakly-isolated mode; (ii) to ensure atomic or order-preserving reads at Minimal Delay imposes to read data from the past (not fresh); (iii) however, order-preserving minimal-delay reads can be fresher than atomic; (iv) reading atomic or order-preserving data at maximal freshness may block reads or writes indefinitely. Our impossibility results hold independently of other features of the database, such as update semantics (totally ordered or not) or data model (structured or unstructured). Guided by these results, we modify an existing protocol to ensure minimal-delay reads (at the cost of freshness) under atomic-visibility and causally-consistent semantics. Our experimental evaluation supports the theoretical results.

This work was published at Middleware 2018 [31].

5.2.2. Co-design and verification of an available file system

Distributed file systems play a vital role in large-scale enterprise services. However, the designer of a distributed file system faces a vexing choice between strong consistency and asynchronous replication. The former supports a standard sequential model by synchronising operations, but is slow and fragile. The latter is highly available and responsive, but exposes users to concurrency anomalies. We describe a rigorous and general approach to navigating this trade-off by leveraging static verification tools that allow to verify different file system designs. We show that common file system operations can run concurrently without synchronisation, while still retaining a semantics reasonably similar to Posix hierarchical structure. The one exception is the “move” operation, for which we prove that, unless synchronised, it will have an anomalous behaviour.

This work was published at VMCAI 2018 [28].

5.3. Resources management in system software

Participants: Michael Damien Carver, Jonathan Lejeune, Pierre Sens, Julien Sopena [correspondent], Gauthier Voron, Francis Laniel.

5.3.1. Multicore schedulers

In collaboration with WHISPER team, we have contributed to an analysis of the impact on application performance of the design and implementation choices made in two widely used open-source schedulers: ULE, the default FreeBSD scheduler, and CFS, the default Linux scheduler. In a paper published at USENIX ATC’18 [24], we compare ULE and CFS in otherwise identical circumstances. This work involves porting ULE to Linux, and using it to schedule all threads that are normally scheduled by CFS. We compare the performance of a large suite of applications on the modified kernel running ULE and on the standard Linux kernel running CFS. The observed performance differences are solely the result of scheduling decisions, and do not reflect differences in other subsystems between FreeBSD and Linux. We found that there is no overall winner. On many workloads the two schedulers perform similarly, but for some workloads there are significant and even surprising differences. ULE may cause starvation, even when executing a single application with identical threads, but this starvation may actually lead to better application performance for some workloads. The more complex load balancing mechanism of CFS reacts more quickly to workload changes, but ULE achieves better load balance in the long run.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

DELYS has a CIFRE contract with Scality SA:

- Dimitrios Vasilas is advised by Marc Shapiro and Brad King. He works on secondary indexing in large-scale storage systems under weak consistency.

DELYS has two CIFRE contracts with Magency SA:

- Damien Carver is advised by Julien Sopena and Sébatien Monnet. He works on designing kernel-level mechanisms that automatically give more memory to the most active containers.
- Lyes Hamidouche is advised by Pierre Sens and Sébatien Monnet. He works on efficient data dissemination among a large number of mobile devices. He defended his thesis in April 2018.

DELYS has two contracts with Orange within the I/O Lab joint laboratory:

- Guillaume Fraysse is advised by Jonathan Lejeune, Julien Sopena, and Pierre Sens. He works on distributed resources allocation in virtual network environments.
- Jonathan Sid-Otmane is advised by Marc Shapiro. He studies the applications of distributed databases to the needs of the telco industry in the context of 5G.

7. Partnerships and Cooperations

7.1. National Initiatives

7.1.1. ANR

7.1.1.1. ESTATE - (2016–2020)

Members: LIP6 (DELYS, project leader), LaBRI (Univ. de Bordeaux); Verimag (Univ. de Grenoble).

Funding: ESTATE is funded by ANR (PRC) for a total of about 544 000 euros, of which 233 376 euros for DELYS.

Objectives: The core of ESTATE consists in laying the foundations of a new algorithmic framework for enabling Autonomic Computing in distributed and highly dynamic systems and networks. We plan to design a model that includes the minimal algorithmic basis allowing the emergence of dynamic distributed systems with self-* capabilities, *e.g.*, self-organization, self-healing, self-configuration, self-management, self-optimization, self-adaptiveness, or self-repair. In order to do this, we consider three main research streams:

(i) building the theoretical foundations of autonomic computing in dynamic systems, (ii) enhancing the safety in some cases by establishing the minimum requirements in terms of amount or type of dynamics to allow some strong safety guarantees, (iii) providing additional formal guarantees by proposing a general framework based on the Coq proof assistant to (semi-)automatically construct certified proofs.

The coordinator of ESTATE is Franck Petit.

7.1.1.2. RainbowFS - (2016–2020)

Members: LIP6 (DELYS, project leader), Scality SA, CNRS-LIG, Télécom Sud-Paris, Université Savoie-Mont-Blanc.

Funding: is funded by ANR (PRC) for a total of 919 534 euros, of which 359 554 euros for DELYS.

Objectives: RainbowFS proposes a “just-right” approach to storage and consistency, for developing distributed, cloud-scale applications. Existing approaches shoehorn the application design to some predefined consistency model, but no single model is appropriate for all uses. Instead, we propose tools to co-design the application and its consistency protocol. Our approach reconciles the conflicting requirements of availability and performance vs. safety: common-case operations are designed to be asynchronous; synchronisation is used only when strictly necessary to satisfy the application’s integrity invariants. Furthermore, we deconstruct classical consistency models into orthogonal primitives that the developer can compose efficiently, and provide a number of tools for quick, efficient and correct cloud-scale deployment and execution. Using this methodology, we will develop an entreprise-grade, highly-scalable file system, exploring the rainbow of possible semantics, and we demonstrate it in a massive experiment.

The coordinator of RainbowFS is Marc Shapiro.

7.1.2. LABEX

7.1.2.1. SMART - (2012–2019)

Members: ISIR (Sorbonne Univ./CNRS), LIP6 (Sorbonne Univ./CNRS), LIB (Sorbonne Univ./INSERM), LJLL (Sorbonne Univ./CNRS), LTCI (Institut Mines-Télécom/CNRS), CHArt-LUTIN (Univ. Paris 8/EPHE), L2E (Sorbonne Univ.), STMS (IRCAM/CNRS).

Funding: Sorbonne Universités, ANR.

Description: The SMART Labex project aims globally to enhancing the quality of life in our digital societies by building the foundational bases for facilitating the inclusion of intelligent artifacts in our daily life for service and assistance. The project addresses underlying scientific questions raised by the development of Human-centered digital systems and artifacts in a comprehensive way. The research program is organized along five axes and DELYS is responsible of the axe "Autonomic Distributed Environments for Mobility."

The project involves a PhD grant of 100 000 euros over 3 years.

7.2. European Initiatives

7.2.1. FP7 & H2020 Projects

7.2.1.1. LightKone

Title: Lightweight Computation for Networks at the Edge

Programm: H2020-ICT-2016-2017

Duration: January 2017 - December 2019

Coordinator: Université Catholique de Louvain

Partners:

Université Catholique de Louvain (Belgium)

Technische Universitaet Kaiserslautern (Germany)

INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciencia (Portugal)

Faculdade de Ciencias E Tecnologiada Universidade Nova de Lisboa (Portugal)

Universitat Politecnica De Catalunya (Spain)

Scality (France)

Gluk Advice B.V. (Netherlands)

Inria contact: Marc Shapiro

The goal of LightKone is to develop a scientifically sound and industrially validated model for doing general-purpose computation on edge networks. An edge network consists of a large set of heterogeneous, loosely coupled computing nodes situated at the logical extreme of a network. Common examples are networks of Internet of Things, mobile devices, personal computers, and points of presence including Mobile Edge Computing. Internet applications are increasingly running on edge networks, to reduce latency, increase scalability, resilience, and security, and permit local decision making. However, today's state of the art, the gossip and peer-to-peer models, give no solution for defining general-purpose computations on edge networks, i.e., computation with shared mutable state. LightKone will solve this problem by combining two recent advances in distributed computing, namely synchronisation-free programming and hybrid gossip algorithms, both of which are successfully used separately in industry. Together, they are a natural combination for edge computing. We will cover edge networks both with and without data center nodes, and applications focused on collaboration, computation, and both. Project results will be new programming models and algorithms that advance scientific understanding, implemented in new industrial applications and a startup company, and evaluated in large-scale realistic settings.

7.3. International Initiatives

7.3.1. Participation in Other International Programs

7.3.1.1. STIC Amsud

Title: PaDMetBio - Parallel and Distributed Metaheuristics for Structural Bioinformatics

International Partners (Institution - Laboratory - Researcher):

Universidade Federal do Rio Grande do Sul (Brazil)- Mãrcio Dorn

Universidad Nacional de San Luis (Argentina) - Verõnica Gil-Costa

Universidad de Santiago de Chile (Chile) - Mario Inostroza-Ponta

Duration: 2017 - 2018

Start year: 2017

Structural bioinformatics deals with problems where the rules that govern the biochemical processes and relations are partially known which makes hard to design efficient computational strategies for these problems. There is a wide range of unanswered questions, which cannot be answered neither by experiments nor by classical modeling and simulation approaches. Specifically, there are several problems that still do not have a computational method that can guarantee a minimum quality of solution. Two of the main challenging problems in Structural Bioinformatics are (1) the three-dimensional (3D) protein structure prediction problem (PSP) and (2) the molecular docking problem for drug design. Predicting the folded structure of a protein only from its amino acid sequence is a challenging problem in mathematical optimization. The challenge arises due to the combinatorial explosion of plausible shapes, where a long amino acid chain ends up in one out of a vast number of 3D conformations. The problem becomes harder when we have proteins with complex topologies, in this case, their predictions may be only possible with significant increases in high-performance computing power. In the case of the molecular docking problem for drug design, we need to predict the preferred orientation of a small drug candidate against a protein molecule. With the increasing availability of molecular biological structures, smarter docking approaches have become necessary. These two problems are classified as NP-Complete or NP-Hard, so there is no current computational approach that can guarantee the best solution for them in a polynomial time. Because of the above, there is the need to build smarter approaches that can deliver good solutions to the problem. In this project, we plan to explore a collaborative work for the design and implementation of population based metaheuristics, like genetic and memetic algorithms. Metaheuristics are one of the most common and powerful techniques used in this case. The main goal of this project is to gather the expertise and current work of researchers in the areas of structural bioinformatics, metaheuristics and parallel and distributed computing, in order to build novel and high quality solutions for these hot research area.

7.3.1.2. Capes-Cofecub

Title: CHOOSING - Cooperation on Hybrid cOmputing clOuds for energy SavING

French Partners: Paris XI (LRI), Regal, LIG, SUPELEC

International Partners (Institution - Laboratory - Researcher):

Universidade de São Paulo - Instituto de Matemática e Estatística - Brazil, Unicamp - Instituto de Computação - Brazil

Duration: 2014–2018

The cloud computing is an important factor for environmentally sustainable development. If, in the one hand, the increasing demand of users drive the creation of large datacenters, in the other hand, cloud computing's “multitenancy” trait allows the reduction of physical hardware and, therefore, the saving of energy. Thus, it is imperative to optimize the energy consumption corresponding to the datacenter's activities. Three elements are crucial on energy consumption of a cloud platform: computation (processing), storage and network infrastructure. Therefore, the aim of this project is to provide different techniques to reduce energy consumption regarding these three elements. Our work mainly focuses on energy saving aspects based on virtualization, i.e., pursuing the idea of the intensive migration of classical storage/processing systems to virtual ones. We will study how different organizations (whose resources are combined as hybrid clouds) can cooperate with each other in order to minimize the energy consumption without the detriment of client requirements or quality of service. Then, we intend to propose efficient algorithmic solutions and design new coordination mechanisms that incentive cloud providers to collaborate.

7.3.1.3. Spanish research ministry project

Title: BFT-DYNASTIE - Byzantine Fault Tolerance: Dynamic Adaptive Services for Partitionable Systems

French Partners: Labri, Irisa, LIP6

International Partners (Institution - Laboratory - Researcher):

University of the Basque Country UPV - Spain, EPFL - LSD - Switzerland, Friedrich-Alexander-Universitat Erlangen-Nurenberg - Deutschland, University of Sydney - Australia

Duration: 2017–2019

The project BFT-DYNASTIE is aimed at extending the model based on the alternation of periods of stable and unstable behavior to all aspects of fault-tolerant distributed systems, including synchrony models, process and communication channel failure models, system membership, node mobility, and network partitioning. The two main and new challenges of this project are: the consideration of the most general and complex to address failure model, known as Byzantine, arbitrary or malicious, which requires qualified majorities and the use of techniques form the security area; and the operation of the system in partitioned mode, which requires adequate reconciliation mechanisms when two partitions merge.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events Organisation

8.1.1.1. General Chair, Scientific Chair

Marc Shapiro, Organiser of Dagstuhl Workshop on “Data Consistency in Distributed Systems: Algorithms, Programs, and Databases” (19-0117), February 2018.

Swan Dubois, PC Track Chair (track A: Theoretical and Practical Aspects of Stabilizing Systems) of the 20th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS 2018).

8.1.1.2. Member of the Organizing Committees

- Marc Shapiro, Steering Committee of Workshop on Principles and Practice of Consistency for distributed Data (PaPoC).
- Franck Petit, Steering Committee of the *International Symposium on Stabilization, Safety, and Security of Distributed Systems* (SSS).
- Pierre Sens, since 2014: Member of Steering Committee of International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD).

8.1.2. Scientific Events Selection

8.1.2.1. Member of Conference Program Committees

Pierre Sens, 28th International Symposium on Software Reliability Engineering (ISSRE 2018), 17th IEEE International Symposium on Network Computing and Applications (NCA 2018), 37th IEEE International Symposium on Reliable Distributed Systems (SRDS 2018), 32rd International Symposium on DIStributed Computing (DISC 2018).

Marc Shapiro, Program Committee of Int. Conf. on Middleware (Middleware 2018).

Marc Shapiro, Program Committee of European Conference on Computer Systems (EuroSys 2019).

Marc Shapiro, Program Committee of Symposium on Principles of Distributed Computing (PODC 2019).

Dimitri Vasilas, Shadow Program Committee of European Conference on Computer Systems (EuroSys 2019).

Franck Petit, 20th Workshop on Advances in Parallel and Distributed Computational Models (APDCM 2018), 24th Conférence d'informatique en Parallélisme, Architecture et Système (COMPAS 2018), 20th Rencontres Francophones sur les Aspects Algorithmiques de Télécommunications (ALGOTEL 2018). Tutorial chairman of Latin-American Symposium on Dependable Computing (LADC 2018).

Luciana Arantes, 17th IEEE International Symposium on Network Computing and Applications (NCA 2018), 14th European Dependable Computing Conference (EDCC 2018), Latin-American Symposium on Dependable Computing (LADC 2018), Conférence d'informatique en Parallélisme, Architecture et Système (COMPAS 2018) .

Swan Dubois, PC member of the ACM Symposium on Principles of Distributed Computing (PODC 2018).

8.1.3. Journal

8.1.3.1. Member of Editorial Boards

Pierre Sens, Associate editor of International Journal of High Performance Computing and Networking (IJHPCN)

Franck Petit, Special Issue on Stabilization, Safety, and Security, Journal on Theory of Computing Systems (ToCS).

Marc Shapiro, Associate Editor for Letters of the IEEE Computer Society (LOCS).

8.1.3.2. Reviewer, Reviewing Activities

Pierre Sens, reviewer Journal of Parallel and Distributed Systems (JPDC), and IEEE Transactions on Parallel and Distributed Systems (TPDS).

Franck Petit, reviewer Journal of the ACM (JACM), Journal of Parallel and Distributed Systems (JPDC), and the International Colloquium on Structural Information and Communication Complexity (SIROCCO 2018).

Luciana Arantes, reviewer Journal of Parallel and Distributed Systems (JPDC), and IEEE Transactions on Parallel and Distributed Systems (TPDS).

Swan Dubois, reviewer Theoretical Computer Science (TCS), ACM Transactions on Computer Systems (TOCS), and The International Symposium on DiStributed Computing (DISC 2018).

8.1.4. Invited Talks

Pierre Sens, *Fault tolerance in dynamic distributed systems*. Invited keynote speaker, "Insights for the Future of Computing", LIG, Grenoble, Avril 2018.

Franck Petit, *Robustness: a New Form of Heredity Motivated by Dynamic Networks*, Invited speaker, 9th Workshop on GRAph Searching, Theory & Applications (GRASTA 2018), Berlin, Germany, Sept. 2018.

Marc Shapiro, *Just-Right Consistency: As available as possible, As consistent as necessary, Correct by design*. Invited speaker, Verification of Distributed Systems workshop, Essaouira, Morocco, May 2018.

Marc Shapiro, *Just-Right Consistency: As available as possible, As consistent as necessary, Correct by design*. Keynote presentation, DotScale, the European Tech Conference on Scalability, Distributed Systems & DevOps, Aubervillers, June 2018.

Marc Shapiro, *Just-Right Consistency*. Keynote presentation, Conf on Advances and Computing and Communication Engineering, Paris, June 2018.

Marc Shapiro, *Just-Right Consistency: As available as possible, Synchronous when necessary, Correct by design*. Invited Keynote talk, I/O Labs annual workshop, Châtillon, France, Oct. 2018.

Marc Shapiro, *Life after consistency*. Workshop of the EuroSys Program Committee, Dec. 2018.

8.1.5. Scientific Expertise

Pierre Sens, Project in Indo-French Centre for the Promotion of Advanded Research

Marc Shapiro, member of Panel PE6 (Computer Science) of European Research Council Starting Grants 2018.

8.1.6. Research Administration

Franck Petit, since 2014: deputy director of the LIP6 laboratory

Pierre Sens, since 2016: Member of Section 6 of the national committee for scientific research CoNRS

Pierre Sens, since 2012: Member of the Executive Committee of Labex SMART, Co-Chair (with F. Petit) of Track 4, Autonomic Distributed Environments for Mobility.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Julien Sopena is Member of “Directoire des formations et de l’insertion professionnelle” of Sorbonne Université, France

Master: Julien Sopena is responsible of Computer Science Master’s degree in Distributed systems and applications (in French, SAR), Sorbonne Universités, France

Master: Luciana Arantes, Swan Dubois, Jonathan Lejeune, Franck Petit, Pierre Sens, Julien Sopena, Advanced distributed algorithms, M2, Sorbonne Université, France

Master: Jonathan Lejeune, Designing Large-Scale Distributed Applications, M2, Sorbonne Université, France

Master: Maxime Lorrillere, Julien Sopena, Linux Kernel Programming, M1, Sorbonne Université, France

Master: Luciana Arantes, Swan Dubois, Jonathan Lejeune, Pierre Sens, Julien Sopena, Operating systems kernel, M1, Sorbonne Université, France

Master: Luciana Arantes, Swan Dubois, Franck Petit, Distributed Algorithms, M1, Sorbonne Université, France

Master: Jonathan Lejeune, Julien Sopena, Client-server distributed systems, M1, Sorbonne Université, France.

Master: Julien Sopena, Marc Shapiro, Ilyas Toumlilt, Francis Laniel. Kernels and virtual machines (*Noyaux et machines virtuelles*, NMV), M2, Sorbonne Université, France.

Licence: Pierre Sens, Luciana Arantes, Julien Sopena, Principles of operating systems, L3, UPMC Sorbonne Université, France

Licence: Swan Dubois, Initiation to operating systems, L3, Sorbonne Université, France

Licence: Swan Dubois, Multi-threaded Programming, L3, Sorbonne Université, France

Licence: Jonathan Lejeune, Oriented-Object Programming, L3, Sorbonne Université, France

Licence: Franck Petit, Advanced C Programming, L2, Sorbonne Université, France

Licence: Swan Dubois, Julien Sopena, Introduction to operating systems, L2, Sorbonne Université, France

Licence: Mesaac Makpangou, C Programming Language, 27 h, L2, Sorbonne Université, France

Ingénieur 4ème année : Marc Shapiro, Introduction aux systèmes d'exploitation, 26 h, M1, Polytech Sorbonne Université, France.

Licence : Philippe Darche (coordinator), Architecture of Internet of Things (IoT), 2 × 32h, L3, Institut Universitaire Technologique (IUT) Paris Descartes, France.

Engineering School: Philippe Darche (coordinator), Solid-State Memories, 4th year, ESIEE, France.

DUT: Philippe Darche (coordinator), Introduction to Computer Systems - Data representation, 60h, Institut Universitaire Technologique (IUT) Paris Descartes, France.

DUT: Philippe Darche (coordinator), Computer Architecture, 32h, Institut Universitaire Technologique (IUT) Paris Descartes, France.

DUT: Philippe Darche (coordinator), Computer Systems Programming, 80h, Institut Universitaire Technologique (IUT) Paris Descartes, France.

8.2.2. Supervision

PhD: Florent Coriat, "Géolocalisation et routage en situation de crise", Dec. 2018, Sorbonne Univ., Anne Fladenmuller (NPA-LIP6) and Luciana Arantes.

PhD: Lyes Hamidouche, "Data replication and data sharing in mobile networks", Sorbonne Univ., CIFRE, Apr. 2018, Sébastien Monnet, Pierre Sens, Dimitri Refauvelet (Magency).

PhD: Denis Jeanneau, "Failure detectors in Dynamic Systems," Sorbonne Université, Dec. 2018, Luciana Arantes, Pierre Sens.

PhD: B Ngom. "FreeCore: a system for indexing document summaries on a Distributed Hash Table (DHT)," Sorbonne Université, Jul. 2018, Mesaac Makpangou.

PhD: Alejandro Z. Tomsic, Sorbonne Univ., defended Apr. 2018, Marc Shapiro. "Computing over widely-replicated data in a hybrid cloud."

PhD: Gauthier Voron, "Big-Os : un OS pour les grands volumes de données,", Sorbonne Univ., Mar. 2018, Gaël Thomas, Pierre Sens.

PhD in progress: João Paulo de Araujo, "L'exécution efficace d'algorithmes distribués dans les réseaux véhiculaires", funded by CNPq (Brésil), since Nov.2015, Pierre Sens and Luciana Arantes.

PhD in progress: Sébastien Bouchard, "Gathering with faulty robots", Sorbonne Univ., since Oct. 2016, Swan Dubois, Franck Petit, Yoann Dieudonné (University of Picardy Jules Verne)

PhD in progress: Marjorie Bournat, "Speculation and Graceful Degradability for Robots in Highly Dynamic Environments", Sorbonne Univ., since Sep. 2015, Swan Dubois, Franck Petit, Yoann Dieudonné (University of Picardy Jules Verne)

PhD in progress: Damien Carver, "HACHE : HorizontAl Cache cHorEgraphy - Toward automatic resizing of shared I/O caches.", Sorbonne Univ., CIFRE, since Jan. 2015, Sébastien Monnet, Pierre Sens, Julien Sopena, Dimitri Refauvelet (Magency).

PhD in progress: Arnaud Favier, "Algorithmes de coordination répartis dans des réseaux dynamiques", Sorbonne Univ., since Sep. 2018, Pierre Sens and Luciana Arantes.

PhD in progress: Saalik Hatia, "Efficient management of memory and storage for CRDTs," Sorbonne Univ., since Oct. 2018. Advised by Marc Shapiro.

CIFRE PhD in progress: Guillaume Fraysse, Orange Lab - Inria, “Ubiquitous Resouces for Service Availability.” Since Jul. 2017, advised by Pierre Sens, Imen Grida Ben Yahia (Orange-Lab) , Jonathan Lejeune, Julien Sopena.

PhD in progress: Francis Laniel, Sorbonne Univ., since Sept. 2017. Advised by Marc Shapiro, Julien Sopena, Jonathan Lejeune. “Vers une utilisation efficace de la mémoire non volatile pour économiser l’énergie.”

PhD in progress: Sreeja Nair, Sorbonne Univ., since April 2018. “Just-Right Consistency for massive gee-replicated storage.” Advised by Marc Shapiro.

CIFRE PhD in progress: Jonathan Sid-Otmane. “Étude des critères de distribution et de l’usage d’une base de données distribuée pour un OS Telco.” Advised by Marc Shapiro, with Sofiane Imadali and Frédéric Martelli, Orange Labs.

PhD in progress: Ilyas Toumlilt, Sorbonne Univ. “Bridging the CAP gap, all the way to the edge.” Advised by Marc Shapiro.

CIFRE PhD in progress: Dimitrios Vasilas, Sorbonne Univ., “Indexing in large-scale storage systems.” Advised by Marc Shapiro, with Brad King, Scality.

8.2.3. Juries

Franck Petit was the reviewer of:

- Alessia Milani, HDR, LaBRI, Bordeaux
- Laurent Feuilloley, PhD, IRIF, Paris

Franck Petit was Chair of

- Florent Coriat, PhD, LIP6, Paris

Pierre Sens was the reviewer of:

- Matthieu Roy, HDR, LAAS, Toulouse
- Thouraya Louati, PhD, Sfax Univ., Tunisie
- Yacine Taleb, PhD, IRSIA, Rennes
- Ye Xia, PhD, LIG, Univ. Grenoble

Pierre Sens was Chair of

- Soraya Zertal, HDR, UVSQ, Versailles
- Nikolaos Georgantas, HDR, Sorbonne Univ. - Inria, Paris
- Alejandro Tomsic, PhD, LIP6, Paris
- Veronica Quintuna Rodriguez, PhD, Sorbonne Univ. - Inria, Paris

Marc Shapiro was a member of the following committees:

- Aurélie Hurault, HdR, ENSEEIHT, Toulouse.
- Soraya Zertal, Comité de suivi doctoral, UVSQ, Versailles.

8.3. Popularization

Jonathan Lejeune and Julien Sopena animated an activity during the Science Festival 2018 at Sorbonne Univ.

9. Bibliography

Major publications by the team in recent years

[1] V. Balegas, N. Preguiça, R. Rodrigues, S. Duarte, C. Ferreira, M. Najafzadeh, M. Shapiro. *Putting Consistency back into Eventual Consistency*, in "Euro. Conf. on Comp. Sys. (EuroSys)", Bordeaux, France, April 2015, pp. 6:1–6:16, https://doi.org/10.1145/2741948.2741972

[2] L. GIDRA, G. THOMAS, J. SOPENA, M. SHAPIRO, N. NGUYEN. *NumaGiC: a garbage collector for big data on big NUMA machines*, in "Int. Conf. on Archi. Support for Prog. Lang. and Systems (ASPLOS)", Istanbul, Turkey, Assoc. for Computing Machinery, March 2015, pp. 661–673, http://dx.doi.org/10.1145/2694344.2694361

[3] A. GOTSMAN, H. YANG, C. FERREIRA, M. NAJAFZADEH, M. SHAPIRO. *'Cause I'm Strong Enough: Reasoning about Consistency Choices in Distributed Systems*, in "Symp. on Principles of Prog. Lang. (POPL)", St. Petersburg, FL, USA, 2016, pp. 371–384, http://dx.doi.org/10.1145/2837614.2837625

[4] M. SHAPIRO, N. PREGUIÇA, C. BAQUERO, M. ZAWIRSKI. *Conflict-free Replicated Data Types*, in "Int. Symp. on Stabilization, Safety, and Security of Dist. Sys. (SSS)", Grenoble, France, X. DÉFAGO, F. PETIT, V. VILLAIN (editors), Lecture Notes in Comp. Sc., Springer-Verlag, October 2011, vol. 6976, pp. 386–400, http://lip6.fr/Marc.Shapiro/papers/CRDTs_SSS-2011.pdf

[5] M. ZAWIRSKI, N. PREGUIÇA, S. DUARTE, A. BIENIUSA, V. BALEGAS, M. SHAPIRO. *Write Fast, Read in the Past: Causal Consistency for Client-side Applications*, in "Int. Conf. on Middleware (MIDDLEWARE)", Vancouver, BC, Canada, ACM/IFIP/Usenix, December 2015, pp. 75–87

Publications of the year

Doctoral Dissertations and Habilitation Theses

[6] F. CORIAT. *Geolocation and communication in post-disaster situations*, Sorbonne Université, Faculté des Sciences et Ingénierie, December 2018, https://tel.archives-ouvertes.fr/tel-01970777

[7] L. HAMIDOUCHE. *Towards efficient dissemination of voluminous data over dense Wi-Fi networks*, Sorbonne Université, Faculté des Sciences et Ingénierie, June 2018, https://hal.inria.fr/tel-01953300

[8] D. JEANNEAU. *Failure Detectors in Dynamic Distributed Systems*, EDITE ; Sorbonne Université, December 2018, https://hal.archives-ouvertes.fr/tel-01951975

[9] B. NGOM. *FreeCore: a system for indexing document summaries on a Distributed Hash Table (DHT)*, Pierre and Marie Curie University, July 2018, https://hal.inria.fr/tel-01921587

[10] A. TOMSIC. *Exploring the design space of highly-available distributed transactions*, Sorbonne Universites, UPMC University of Paris 6, April 2018, https://hal.archives-ouvertes.fr/tel-01956321

[11] G. VORON. *Efficient Virtualization of NUMA Architectures*, Sorbonne Université, Faculté des Sciences et Ingénierie, March 2018, https://hal.archives-ouvertes.fr/tel-01947560

Articles in International Peer-Reviewed Journals

[12] Z. AL-SHARA, F. ALVARES, H. BRUNELIERE, J. LEJEUNE, C. PRUD'HOMME, T. LEDOUX. *CoMe4ACloud: An End-to-End Framework for Autonomic Cloud Systems*, in "Future Generation Computer Systems", September 2018, vol. 86, pp. 339-354 [*DOI :* 10.1016/J.FUTURE.2018.03.039], https://hal.archives-ouvertes.fr/hal-01762716

[13] L. BLIN, F. BOUBEKEUR, S. DUBOIS. *A Self-Stabilizing Memory Efficient Algorithm for the Minimum Diameter Spanning Tree under an Omnipotent Daemon*, in "Journal of Parallel and Distributed Computing", July 2018, vol. 117, pp. 50-62 [*DOI :* 10.1016/J.JPDC.2018.02.007], https://hal.inria.fr/hal-01966265

[14] S. BOUCHARD, Y. DIEUDONNÉ, F. PETIT, A. PELC. *On Deterministic Rendezvous at a Node of Agents with Arbitrary Velocities*, in "Information Processing Letters", January 2018, vol. 133, pp. 39 - 43 [*DOI :* 10.1016/J.IPL.2018.01.003], https://hal.archives-ouvertes.fr/hal-01701786

[15] A. GRACIELA DE MORAES ROSSETTO, C. GEYER, L. ARANTES, P. SENS. *Impact FD: An Unreliable Failure Detector Based on Process Relevance and Confidence in the System*, in "The Computer Journal", 2018, https://hal.inria.fr/hal-01793311

[16] L. A. RODRIGUES, E. P. DUARTE JR., L. ARANTES. *A distributed k-mutual exclusion algorithm based on autonomic spanning trees*, in "Journal of Parallel and Distributed Computing", May 2018, vol. 115, pp. 41-55 [*DOI :* 10.1016/J.JPDC.2018.01.008], https://hal.inria.fr/hal-01965673

[17] J. P. DE ARAUJO, L. ARANTES, E. P. DUARTE JR., L. A. RODRIGUES, P. SENS. *VCube-PS: A causal broadcast topic-based publish/subscribe system*, in "Journal of Parallel and Distributed Computing", November 2018 [*DOI :* 10.1016/J.JPDC.2018.10.011], https://hal.inria.fr/hal-01925856

Articles in Non Peer-Reviewed Journals

[18] A. BIENIUSA, A. GOTSMAN, B. KEMME, M. SHAPIRO. *Data Consistency in Distributed Systems: Algorithms, Programs, and Databases*, in "Dagstuhl Reports", July 2018, vol. 8, n^{o} 2, pp. 101-121 [*DOI :* 10.4230/DAGREP.8.2.101], https://hal.inria.fr/hal-01848384

International Conferences with Proceedings

[19] L. ARANTES, E. BAMPIS, A. KONONOV, M. LETSIOS, G. LUCARELLI, P. SENS. *Scheduling under Uncertainty: A Query-based Approach*, in "IJCAI 2018 - 27th International Joint Conference on Artificial Intelligence", Stockholm, Sweden, July 2018, https://hal.inria.fr/hal-01924648

[20] S. BOUCHARD, M. BOURNAT, Y. DIEUDONNÉ, S. DUBOIS, F. PETIT. *Approche asynchrone dans le plan : un algorithme déterministe polynomial*, in "ALGOTEL 2018 - 20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Roscoff, France, May 2018, https://hal.archives-ouvertes.fr/hal-01782388

[21] S. BOUCHARD, Y. DIEUDONNÉ, B. DUCOURTHIAL. *Rassemblement byzantin dans les réseaux*, in "20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications(ALGOTEL 2018)", Roscoff, France, May 2018, https://hal.archives-ouvertes.fr/hal-01782387

[22] S. BOUCHARD, Y. DIEUDONNÉ, A. LAMANI. *Byzantine Gathering in Polynomial Time*, in "45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)", Prague, Czech Republic, July 2018 [*DOI :* 10.4230/LIPICS.ICALP.2018.147], https://hal.archives-ouvertes.fr/hal-01965743

[23] S. BOUCHARD, Y. DIEUDONNÉ, A. PELC, F. PETIT. *Deterministic Treasure Hunt in the Plane with Angular Hints*, in "29th International Symposium on Algorithms and Computation, ISAAC 2018", Jiaoxi Township, Taiwan, W.-L. HSU, D.-T. LEE, C.-S. LIAO (editors), Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, December 2018, vol. 123, pp. 48:1–48:13 [*DOI :* 10.4230/LIPICS.ISAAC.2018.48], https://hal.sorbonne-universite.fr/hal-01970990

[24] J. BOURON, S. CHEVALLEY, B. LEPERS, W. ZWAENEPOEL, R. GOUICEM, J. LAWALL, G. MULLER, J. SOPENA. *The Battle of the Schedulers: FreeBSD ULE vs. Linux CFS*, in "2018 USENIX Annual Technical Conference", Boston, MA, United States, July 2018, https://hal.inria.fr/hal-01853267

[25] H. BRUNELIERE, Z. AL-SHARA, F. ALVARES, J. LEJEUNE, T. LEDOUX. *A Model-based Architecture for Autonomic and Heterogeneous Cloud Systems*, in "CLOSER 2018 - 8h International Conference on Cloud Computing and Services Science", Funchal, Portugal, March 2018, vol. 1, pp. 201-212, Best Paper Award [*DOI :* 10.5220/0006773002010212], https://hal.archives-ouvertes.fr/hal-01705248

[26] G. FRAYSSE, I. GRIDA BEN YAHIA, J. LEJEUNE, P. SENS, J. SOPENA. *Towards multi-SDN services: Dangers of concurrent resource allocation from multiple providers*, in "21st Conference on Innovation in Clouds, Internet and Networks (ICIN 2018)", Paris, France, February 2018, https://hal.inria.fr/hal-01793636

[27] E. MAUFFRET, D. JEANNEAU, L. ARANTES, P. SENS. *The Weakest Failure Detector to Solve the Mutual Exclusion Problem in an Unknown Dynamic Environment*, in "20th International Conference on Distributed Computing and Networking (ICDCN 2019)", Bangalore, India, January 2019, Extended version: https://hal.archives-ouvertes.fr/hal-01661127v3, https://hal.archives-ouvertes.fr/hal-01929224

[28] M. NAJAFZADEH, M. SHAPIRO, P. EUGSTER. *Co-Design and Verification of an Available File System*, in "VMCAI 2018 - International Conference on Verification, Model Checking, and Abstract Interpretation", Los Angeles, CA, United States, I. DILLIG, J. PALSBERG (editors), Lecture Notes in Computer Science, Springer, January 2018, vol. 10747, pp. 358-381 [*DOI :* 10.1007/978-3-319-73721-8_17], https://hal.inria.fr/hal-01696263

[29] L. A. RODRIGUES, E. P. DUARTE JR., J. P. DE ARAUJO, L. ARANTES, P. SENS. *Bundling Messages to Reduce the Cost of Tree-Based Broadcast Algorithms*, in "LADC 2018 - 8th Latin-American Symposium on Dependable Computing", Foz do Iguaçu, Brazil, October 2018, https://hal.inria.fr/hal-01959395

[30] M. SHAPIRO, A. BIENIUSA, P. ZELLER, G. PETRI. *Ensuring referential integrity under causal consistency*, in "PaPoC 2018 - 5th Workshop on Principles and Practice of Consistency for Distributed Data", Porto, Portugal, April 2018, https://arxiv.org/abs/1803.03482 , https://hal.inria.fr/hal-01727207

[31] A. Z. TOMSIC, M. BRAVO, M. SHAPIRO. *Distributed transactional reads: the strong, the quick, the fresh & the impossible*, in "2018 ACM/IFIP/USENIX International Middleware Conference", Rennes, France, Proceedings of 2018 ACM/IFIP/USENIX International Middleware Conference, ACM, December 2018, 14 p. , https://arxiv.org/abs/1810.01698 [*DOI :* 10.1145/3274808.3274818], https://hal.inria.fr/hal-01876456

[32] D. VASILAS, M. SHAPIRO, B. KING. *A Modular Design for Geo-Distributed Querying: Work in Progress Report*, in "PaPoC 2018 - 5th Workshop on Principles and Practice of Consistency for Distributed Data", Porto, Portugal, April 2018, pp. 1-8, https://arxiv.org/abs/1803.04141 , https://hal.inria.fr/hal-01728712

[33] J. P. DE ARAUJO, L. ARANTES, E. P. DUARTE JR., L. A. RODRIGUES, P. SENS. *A Communication-Efficient Causal Broadcast Protocol*, in "ICPP 2018 - 47th International Conference on Parallel Processing", Eugene, Oregon, United States, August 2018, https://hal.inria.fr/hal-01924741

Conferences without Proceedings

[34] K. ALTISEN, S. DEVISMES, A. DURAND, F. PETIT. *Stabilisation progressive*, in "ALGOTEL 2018 - 20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Roscoff, France, May 2018, https://hal.archives-ouvertes.fr/hal-01779963

[35] F. CORIAT, A. FLADENMULLER, L. ARANTES. *Architecture de collecte pour la géolocalisation en situation de crise : évaluation comparative*, in "ComPAS 2018 - Conférence d'informatique en Parallélisme, Architecture et Système", Toulouse, France, July 2018, https://hal.sorbonne-universite.fr/hal-01877942

Research Reports

[36] M. BOURNAT, S. DUBOIS, F. PETIT. *Gracefully Degrading Gathering in Dynamic Rings*, LIP6, Sorbonne Université, CNRS, UMR 7606 ; DELYS ; Inria, May 2018, https://arxiv.org/abs/1805.05137 , https://hal.archives-ouvertes.fr/hal-01790554

[37] E. MAUFFRET, D. JEANNEAU, L. ARANTES, P. SENS. *The Weakest Failure Detector to Solve the Mutual Exclusion Problem in an Unknown Dynamic Environment*, LISTIC ; Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, October 2018 [*DOI :* 10.4230/LIPICS], https://hal.archives-ouvertes.fr/hal-01661127

[38] S. S. NAIR, M. SHAPIRO. *Improving the "Correct Eventual Consistency" Tool*, Sorbonne Université, July 2018, n° RR-9191, https://arxiv.org/abs/1807.06431 , https://hal.inria.fr/hal-01832888

[39] M. SHAPIRO, A. BIENIUSA, N. PREGUIÇA, V. BALEGAS, C. MEIKLEJOHN. *Just-Right Consistency: reconciling availability and safety*, Inria Paris ; UPMC - Paris 6 Sorbonne Universités ; Tech. U. Kaiserslautern ; U. Nova de Lisboa ; U. Catholique de Louvain, January 2018, n° RR-9145, pp. 1-15, https://arxiv.org/abs/1801.06340 , https://hal.inria.fr/hal-01685945

Scientific Popularization

[40] N. PREGUIÇA, C. BAQUERO, M. SHAPIRO. *Conflict-free Replicated Data Types (CRDTs)*, in "Encyclopedia of Big Data Technologies", Encyclopedia of Big Data Technologies, Springer International Publishing, April 2018, https://arxiv.org/abs/1805.06358 [*DOI :* 10.1007/978-3-319-63962-8_185-1], https://hal.inria.fr/hal-01793586

[41] M. SHAPIRO, P. SUTRA. *Database Consistency Models*, in "Encyclopedia of Big Data Technologies", S. SAK, A. ZOMAYA (editors), Springer, February 2018, https://arxiv.org/abs/1804.00914 [*DOI :* 10.1007/978-3-319-63962-8_203-1], https://hal.inria.fr/hal-01756780

Other Publications

[42] B. NGOM, M. MAKPANGOU, S. NDIAYE. *SPT – Summary Prefix Tree: An over DHT Indexing Data Structure for Efficient Superset Search*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01757074