IN PARTNERSHIP WITH:
CNRS

Université Pierre et Marie Curie (Paris 6)

Activity Report 2017

Project-Team REGAL

Large-Scale Distributed Systems and Applications

IN COLLABORATION WITH: Laboratoire d'informatique de Paris 6 (LIP6)

RESEARCH CENTER
Paris

THEME
Distributed Systems and middleware

Table of contents

Project-Team REGAL

Creation of the Project-Team: 2005 July 01, end of the Project-Team: 2017 December 31

Keywords:

Computer Science and Digital Science:

A1.1.1. - Multicore, Manycore
A1.1.6. - Cloud
A1.1.7. - Peer to peer
A1.1.9. - Fault tolerant systems
A1.1.13. - Virtualization
A1.3. - Distributed Systems
A1.6. - Green Computing
A2.6. - Infrastructure software
A2.6.1. - Operating systems
A2.6.2. - Middleware
A2.6.3. - Virtual machines
A3.1.3. - Distributed data
A3.1.8. - Big data (production, storage, transfer)
A7.1. - Algorithms

Other Research Topics and Application Domains:

B4.5. - Energy consumption
B6.4. - Internet of things
B6.6. - Embedded systems
B8.2. - Connected city
B9.2.3. - Video games
B9.4.1. - Computer science

1. Personnel

Research Scientists

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

Faculty Members

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

Post-Doctoral Fellow

Paolo Viotti [Univ Pierre et Marie Curie]

PhD Students

Sébastien Bouchard [Inria]
Marjorie Bournat [Univ Pierre et Marie Curie]
Damien Carver [Magency]
João Paulo de Araujo [Univ Pierre et Marie Curie]
Guillaume Fraysse [IO Lab, Orange Lab]
Lyes Hamidouche [Magency]
Denis Jeanneau [Univ Pierre et Marie Curie]
Francis Laniel [Univ Pierre et Marie Curie]
Vinh Tao Thanh [Scality, until Dec 2017]
Alejandro Tomsic [Inria]
Ilyas Toumlilt [Univ Pierre et Marie Curie]
Dimitrios Vasilas [Scality]
Gauthier Voron [Univ Pierre et Marie Curie]

Technical staff
Sreeja Nair [Univ Pierre et Marie Curie]

Administrative Assistant
Nelly Maloisel [Inria]

External Collaborator
Sébastien Monnet [Université Savoie Mont-Blanc]

2. Overall Objectives

2.1. Overall Objectives

The research of the Regal team addresses the theory and practice of *Computer Systems,* including multicore computers, clusters, networks, peer-to-peer systems, cloud computing systems, and other communicating entities such as swarms of robots. It addresses the challenges of communicating, sharing information, and computing correctly 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.

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

3. Research Program

3.1. Research rationale

The research of Regal addresses both theoretical and practical issues of *Computer Systems*, i.e., its goal is a dual expertise in theoretical and experimental research. Our approach is a "virtuous cycle" of algorithm design triggered by issues with real systems, which we prove correct and evaluate theoretically, and then eventually implement and test experimentally.

Regal's major challenges comprise communication, sharing of information, and correct execution in large-scale and/or highly dynamic computer systems. While Regal's historically focused in static distributed systems, since some years ago we have covered a larger spectrum of distributed computer systems: multicore computers, clusters, mobile networks, peer-to-peer systems, cloud computing systems, and other communicating entities such as swarms of robots. This holistic approach allows the handling of related problems at different levels. Among such problems we can highlight communication between cores, consensus, fault detection, scalability, search and diffusion of information, allocation resource, replication and consistency of shared data, dynamic content distribution, and multi-core concurrent algorithms.

Computer Systems is a rapidly evolving domain, with strong interactions with industry and modern computer systems, which are increasingly distributed. Ensuring persistence, availability, and consistency of data in a distributed setting is a major requirement: the system must remain correct despite slow networks, disconnection, crashes, failures, churn, and attacks. Easiness of use, performance, and efficiency are equally fundamental. However, these requirements are somewhat conflicting, and there are many algorithmic and engineering trade-offs, which often depend on specific workloads or usage scenarios. At the same time, years of research in distributed systems are now coming to fruition, and are being used by millions of users of web systems, peer-to-peer systems, gaming and social applications, or cloud computing. These new usages bring new challenges of extreme scalability and adaptation to dynamically-changing conditions, where knowledge of the system state might only be partial and incomplete. Therefore, the scientific challenges of the distributed computing systems listed above are subject to additional trade-offs which include scalability, fault tolerance, dynamics, and virtualization of physical infrastructure. 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.

In in particular, Regal focuses on three key challenges:

- the adaptation of algorithms to the new dynamics of distributed systems;
- data management on extreme large configurations;
- the adaptation of execution support to new multi-core architectures.

We should emphasize that these challenges are complementary: the two first challenges aim at building new distributed algorithms and strategies for large and dynamic distributed configurations whereas the last one focusses on the scalability of internal OS mechanisms.

4. New Software and Platforms

4.1. Antidote

KEYWORDS: Distributed computing - Distributed Data Management - Cloud storage - Large scale
FUNCTIONAL DESCRIPTION: Antidote is the flexible cloud database platform currently under development in the SyncFree and LightKone European projects. Antidote aims to be both a research platform for studying replication and consistency at the large scale, and an instrument for exploiting research results. The platform supports replication of CRDTs, in and between sharded (partitioned) data centres (DCs). The current stable version supports strong transactional consistency inside a DC, and causal transactional consistency between DCs. Ongoing research includes support for explicit consistency , for elastic version management, for adaptive replication, for partial replication, and for reconfigurable sharding.

- Participants: Marc Shapiro, Paolo Viotti, Alejandro Tomsic, Ilyas Toumlilt and Dimitrios Vasilas
- Partners: Université Catholique de Louvain (UCL), Louvain-la-Neuve,Belgium - Universidade NOVA de Lisboa, Portugal - Technische Universität Kaiserslautern (UniKL), Allemagne
- Contact: Marc Shapiro
- Publications: Bringing the cloud closer to users - Write Fast, Read in the Past: Causal Consistency for Client-side Applications - Extending Eventually Consistent Cloud Databases for Enforcing Numeric Invariants - Designing a causally consistent protocol for geo-distributed partial replication - Towards Fast Invariant Preservation in Geo-replicated Systems - Putting Consistency back into Eventual Consistency - The Case for Fast and Invariant-Preserving Geo-Replication - Improving the scalability of geo-replication with reservations - Conflict-free Replicated Data Types - An encounter with Marc Shapiro and his SyncFree European project - PhysiCS-NMSI: efficient consistent snapshots for scalable snapshot isolation - Geo-Replication: Fast If Possible, Consistent If Necessary - Cure: Strong semantics meets high availability and low latency - Cure: Strong semantics meets high availability and low latency

4.2. CISE Tool

KEYWORDS: Distributed Applications - Program verification
FUNCTIONAL DESCRIPTION: Static analysis of the model of a distributed application, to prove (under the assumption of causal consistency) whether the invariants of the application are always satisfied, and to provide a counterexample if not.

- Participants: Sreeja Nair and Marc Shapiro
- Contact: Marc Shapiro
- Publications: Evaluation of the CEC (Correct Eventual Consistency) Tool - The CISE Tool: Proving Weakly-Consistent Applications Correct - The CISE Tool: Proving Weakly-Consistent Applications Correct - CISE Safety Tool - 'Cause I'm Strong Enough: Reasoning about Consistency Choices in Distributed Systems - Putting Consistency back into Eventual Consistency

4.3. PUMA

Puma: pooling unused memory in virtual machines
KEYWORDS: Virtualization - Operating system - Distributed systems - Linux kernel
FUNCTIONAL DESCRIPTION: PUMA is a system that is based on a kernel-level remote caching mechanism that provides the ability to pool VMs memory at the scale of a data center. An important property while lending memory to another VM, is the ability to quickly retrieve memory in case of need. Our approach aims at lending memory only for clean cache pages: in case of need, the VM which lent the memory can retrieve it easily. We use the system page cache to store remote pages such that: (i) if local processes allocate memory the borrowed memory can be retrieved immediately, and (ii) if they need cache the remote pages have a lower priority than the local ones.

- Participants: Maxime Lorrillere, Julien Sopena and Pierre Sens
- Partner: LIP6
- Contact: Julien Sopena
- Publications: Conception et évaluation d'un système de cache réparti adapté aux environnements virtualisés - Puma: pooling unused memory in virtual machines for I/O intensive applications
- URL: https://github.com/mlorrillere/puma

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, 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 2017, we obtained the following results.

5.1.1. Algorithms for Dynamic and Large Systems

In [32] we propose VCube-PS, a new topic-based Publish/Subscribe system built on the top of a virtual hypercube like topology. Membership information and published messages to subscribers (members) of a topic group are broadcast over dynamically built spanning trees rooted at the message's source. For a given topic, delivery of published messages respects causal order. Performance results of experiments conducted on the PeerSim simulator confirm the efficiency of VCube-PS in terms of scalability, latency, number, and size of messages when compared to a single rooted, not dynamically, tree built approach.

We also explore in [20] scheduling challenges in providing probabilistic Byzantine fault tolerance in a hybrid cloud environment, consisting of nodes with varying reliability levels, compute power, and monetary cost. In this context, the probabilistic Byzantine fault tolerance guarantee refers to the confidence level that the result of a given computation is correct despite potential Byzantine failures. We formally define a family of such scheduling problems distinguished by whether they insist on meeting a given latency limit and trying to optimize the monetary budget or vice versa. For the case where the latency bound is a restriction and the budget should be optimized, we propose several heuristic protocols and compare between them using extensive simulations.

In [27], we a new resource reservation protocol in the context of delay-sensitive rescue mobile networks. The search for service providers (e.g., ambulance, fire truck, etc.) after a disaster, must take place within a short time. Therefore, service discovery protocol which looks for providers that can attend victims, respecting time constraints, is crucial. In such a situation, a commonly solution for ensuring network connectivity between victims and providers is ad hoc networks (MANET), composed by battery-operated mobile nodes of persons (victims or not). Using message aggregations techniques, we propose an new reservation protocol aiming at reducing the number of messages over the network and, therefore, node's battery consumption

5.1.2. Self-Stabilization

Self-stabilization is a generic paradigm to tolerate transient faults (*i.e.*, faults of finite duration) in distributed systems. In [14], we propose a silent self-stabilizing leader election algorithm for bidirectional arbitrary connected identified networks. This algorithm is written in the locally shared memory model under the distributed unfair daemon. It requires no global knowledge on the network. Its stabilization time is in $\Theta(n^3)$ steps in the worst case, where n is the number of processes. Its memory requirement is asymptotically optimal, *i.e.*, $\Theta(\log n)$ bits per processes. Its round complexity is of the same order of magnitude — *i.e.*, $\Theta(n)$ rounds — as the best existing algorithms designed with similar settings. To the best of our knowledge, this is the first self-stabilizing leader election algorithm for arbitrary identified networks that is proven to achieve a stabilization time polynomial in steps. By contrast, we show that the previous best existing algorithms designed with similar settings stabilize in a non polynomial number of steps in the worst case.

5.1.3. Mobile Agents

In [21], we consider systems made of autonomous mobile robots evolving in highly dynamic discrete environment *i.e.*, graphs where edges may appear and disappear unpredictably without any recurrence, stability, nor periodicity assumption. Robots are uniform (they execute the same algorithm), they are anonymous (they are devoid of any observable ID), they have no means allowing them to communicate together, they share no common sense of direction, and they have no global knowledge related to the size of the environment. However, each of them is endowed with persistent memory and is able to detect whether it stands alone at its current location. A highly dynamic environment is modeled by a graph such that its topology keeps continuously changing over time. In this paper, we consider only dynamic graphs in which nodes are anonymous, each of them is infinitely often reachable from any other one, and such that its underlying graph (*i.e.*, the static graph made of the same set of nodes and that includes all edges that are present at least once over time) forms a ring of arbitrary size.

In this context, we consider the fundamental problem of *perpetual exploration*: each node is required to be infinitely often visited by a robot. This paper analyzes the computability of this problem in (fully) synchronous settings, *i.e.*, we study the deterministic solvability of the problem with respect to the number of robots. We

provide three algorithms and two impossibility results that characterize, for any ring size, the necessary and sufficient number of robots to perform perpetual exploration of highly dynamic rings.

5.1.4. Approach in the Plane

In [35] we study the task of *approach* of two mobile agents having the same limited range of vision and moving asynchronously in the plane. This task consists in getting them in finite time within each other's range of vision. The agents execute the same deterministic algorithm and are assumed to have a compass showing the cardinal directions as well as a unit measure. On the other hand, they do not share any global coordinates system (like GPS), cannot communicate and have distinct labels. Each agent knows its label but does not know the label of the other agent or the initial position of the other agent relative to its own. The route of an agent is a sequence of segments that are subsequently traversed in order to achieve approach. For each agent, the computation of its route depends only on its algorithm and its label. An adversary chooses the initial positions of both agents in the plane and controls the way each of them moves along every segment of the routes, in particular by arbitrarily varying the speeds of the agents. Roughly speaking, the goal of the adversary is to prevent the agents from solving the task, or at least to ensure that the agents have covered as much distance as possible before seeing each other. A deterministic approach algorithm is a deterministic algorithm that always allows two agents with any distinct labels to solve the task of approach regardless of the choices and the behavior of the adversary. The cost of a complete execution of an approach algorithm is the length of both parts of route travelled by the agents until approach is completed.

Let Δ and l be the initial distance separating the agents and the length of (the binary representation of) the shortest label, respectively. *Assuming that Δ and l are unknown to both agents, does there exist a deterministic approach algorithm whose cost is polynomial in Δ and l?*

Actually the problem of approach in the plane reduces to the network problem of rendezvous in an infinite oriented grid, which consists in ensuring that both agents end up meeting at the same time at a node or on an edge of the grid. By designing such a rendezvous algorithm with appropriate properties, as we do in this paper, we provide a positive answer to the above question.

Our result turns out to be an important step forward from a computational point of view, as the other algorithms allowing to solve the same problem either have an exponential cost in the initial separating distance and in the labels of the agents, or require each agent to know its starting position in a global system of coordinates, or only work under a much less powerful adversary.

5.2. Large scale data distribution

Participants: Mesaac Makpangou, Sébastien Monnet, Pierre Sens, Marc Shapiro, Paolo Viotti, Sreeja Nair, Ilyas Toumlilt, Alejandro Tomsic, Dimitrios Vasilas.

5.2.1. Data placement and searches over large distributed storage

Distributed storage systems such as Hadoop File System or Google File System (GFS) ensure data availability and durability using replication. Persistence is achieved by replicating the same data block on several nodes, and ensuring that a minimum number of copies are available on the system at any time. Whenever the contents of a node are lost, for instance due to a hard disk crash, the system regenerates the data blocks stored before the failure by transferring them from the remaining replicas. In [33] we focused on the analysis of the efficiency of replication mechanism that determines the location of the copies of a given file at some server. The variability of the loads of the nodes of the network is investigated for several policies. Three replication mechanisms are tested against simulations in the context of a real implementation of a such a system: Random, Least Loaded and Power of Choice. The simulations show that some of these policies may lead to quite unbalanced situations. It is shown in this paper that a simple variant of a power of choice type algorithm has a striking effect on the loads of the nodes. Mathematical models are introduced and investigated to explain this interesting phenomenon. The analysis of these systems turns out to be quite complicated mainly because of the large dimensionality of the state spaces involved. Our study relies on probabilistic methods, mean-field analysis, to analyze the asymptotic behavior of an arbitrary node of the network when the total number of nodes gets large.

In the summary prefix tree (SPT), a trie data structure that supports efficient superset searches over DHT. Each document is summarized by a Bloom filter which is then used by SPT to index this document. SPT implements an hybrid lookup procedure that is well-adapted to sparse indexing keys such as Bloom filters. It also proposes a mapping function that permits to mitigate the impact of the skewness of SPT due to the sparsity of Bloom filters, especially when they contain only few words. To perform efficient superset searches, SPT maintains on each node a local view of the global tree. The main contributions are the following. First, the approximation of the superset relationship among keyword-sets by the descendance relationship among Bloom filters. Second, the use of a summary prefix tree (SPT), a trie indexing data structure, for keyword-based search over DHT. Third, an hybrid lookup procedure which exploits the sparsity of Bloom filters to offer good performances. Finally, an algorithm that exploits SPT to efficiently find descriptions that are supersets of query keywords.

5.2.2. Just-Right Consistency

Consistency is a major concern in the design of distributed applications, but the topic is still not well understood. It is clear that no single consistency model is appropriate for all applications, but how do developers find their way in the maze of models and the inherent trade-offs between correctness and availability? The Just-Right Consistency approach presented here offers some guidance. First, we classify the safety patterns that are of interest to maintain application correctness. Second, we show how two of these patterns are "AP-compatible" and can be guaranteed without impacting availability, thanks to an appropriate data model and consistency model. Then we address the last, "CAP-sensitive" pattern. In a restricted but common case it can be maintained efficiently in a mostly-available way. In the general case, we exhibit a static analysis logic and tool which ensures just enough synchronisation to maintain the invariant, and availability otherwise.

In summary, instead of pre-defining a consistency model and shoe-horning the application to fit it, and instead of making the application developer compensate for the imperfections of the data store in an *ad-hoc* way, we have a provably correct approach to tailoring consistency to the specific application requirements. This approach is supported by several artefacts developed by Regal and collaborators: Conflict-Free Replicated Data Types (CRDTs), the Antidote cloud database, and the CISE verification tool.

This paper is under submission.

5.3. Memory management in system software

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

Recent years have seen the increasingly widespread use of **multicore** architectures and **virtualized environments**. This development has an impact on all parts of the system software. Virtual machine (VM) technology offers both isolation and flexibility but has side effects such as fragmentation of the physical resources, including memory. This fragmentation reduces the amount of available memory a VM can use. Many recent works study that a NUMA (Non Uniform Memory Access) architecture, common in large multi-core processors, highly impacts application performance. We focus on improving the memory and cache management in various virtualized environments such as Xen hypervisor or linux-containers targeting big data applications on multicore architectures.

While virtualization only introduces a small overhead on machines with few cores, this is not the case on larger ones. Most of the overhead on the latter machines is caused by the NUMA architecture they are using. In order to reduce this overhead, in [34] we show how NUMA placement heuristics can be implemented inside Xen. With an evaluation of 29 applications on a 48-core machine, we show that the NUMA placement heuristics can multiply the performance of 9 applications by more than 2.

We also study the memory arbitration between containers. In the Damien Carver's PhD thesis, we are designing ACDC [23] (Advanced Consolidation for Dynamic Containers), a kernel-level mechanisms that automatically provides more memory to the most active containers.

In the Francis Laniel's PhD thesis, we study a new architecture using Non Volatile RAM NVRAM. Although NVRAM are slower than classical RAM, they have better energetic features. We investigate solutions where RAM and NVRAM coexist in order to balance the energy consumption and performance according to the needs of the system.

6. Bilateral Contracts and Grants with Industry

6.1. Bilateral Contracts with Industry

Regal has two CIFRE contracts with Scality SA:

- Vinh Tao is advised by Marc Shapiro and Vianney Rancurel. He works on highly available geo-replicated file systems, building on CRDT technology. He defended his thesis in December 2017.
- Dimitrios Vasilas is advised by Marc Shapiro and Brad King. He works on secondary indexing in large-scale storage systems under weak consistency.

Regal 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.

Regal has one contract 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.

7. Partnerships and Cooperations

7.1. National Initiatives

7.1.1. Labex SMART - (2012–2019)

Members: ISIR (UPMC/CNRS), LIP6 (UPMC/CNRS), LIB (UPMC/INSERM), LJLL (UPMC/CNRS), LTCI (Institut Mines-Télécom/CNRS), CHArt-LUTIN (Univ. Paris 8/EPHE), L2E (UPMC), 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 Regal is responsible of the axe "Autonomic Distributed Environments for Mobility."

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

7.1.2. ESTATE - (2016–2020)

Members: LIP6 (Regal, 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 Regal.

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.

7.1.3. RainbowFS - (2016–2020)

Members: LIP6 (Regal, 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 Regal.

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.

7.2. 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. CNRS-Inria-FAP's

Title: Autonomic and Scalable Algorithms for Building Resilient Distributed Systems

International Partner (Institution - Laboratory - Researcher):

Universida de Federal do Paraná (UFPR), Brazil, Prof. Elias Duarte

Duration: 2015–2017

In the context of autonomic computing systems that detect and diagnose problems, self-adapting themselves, the VCube (Virtual Cube), proposed by Prof. Elias Duarte , is a distributed diagnosis algorithm that organizes the system nodes on a virtual hypercube topology. VCube has logarithmic properties: when all nodes are fault-free, processes are virtually connected to form a perfect hypercube; as soon as one or more failures are detected, links are automatically reconnected to remove the faulty nodes and the resulting topology, connecting only fault-free nodes, keeps the logarithmic properties. The goal of this project is to exploit the autonomic and logarithmic properties of the VCube by proposing self-adapting and self-configurable services.

7.3.1.3. 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.

7.3.1.4. 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.

7.4. International Research Visitors

7.4.1. Visits of International Scientists

7.4.1.1. Internships

Ajay Singh of Indian Institute Of Technology Hyderabad, India, was invited for a six-month internship, on data structures for concurrency and persistent memory. This work is published at the HiPC SRS 2017 workshop [43].

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, Co-organizer (with Arnaud Casteigts, University of Bordeaux) of the Second Workshop on Computing in Dynamic Networks, in conjunction with DISC'17, Vienna, Austria, October 20th, 2017.

8.1.2. Scientific Events Selection

8.1.2.1. Member of the Conference Program Committees

Pierre Sens, 28th International Symposium on Software Reliability Engineering (ISSRE'2017), IEEE 46th International Conference on Parallel Processing (ICPP'2017), 16th IEEE International Symposium on Network Computing and Applications (NCA 2017).

Marc Shapiro, Steering Committee of Int. Conf. on Principles of Distributed Systems (OPODIS).

Marc Shapiro, Steering Committed of Workshop on Principles and Practice of Consistency for Distr. Data (PaPoC).

Swan Dubois, 19th Workshop on Advances in Parallel and Distributed Computational Models (APDCM'2017), 5th International Symposium on Computing and Networking (CANDAR'2017).

Luciana Arantes, 16th IEEE International Symposium on Network Computing and Applications (NCA 2017), 13th European Dependable Computing Conference (EDCC 2017), 17th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS 2017)

8.1.2.2. Reviewer

- Marc Shapiro, reviewer for Int. Conf. on Middleware (MIDDLEWARE 2017).
- Swan Dubois, 31st International Symposium on Distributed Computing (DISC'2017), 24th International Colloquium on Structural Information and Communication Complexity (SIROCCO'2017).
- Luciana Arantes, ACM Symposium on Principles of Distributed Computing (PODC 2017), 23rd International European Conference on Parallel and Distributed Computing (Euro-Par 2017).

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

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

Lucian Arantes, Special Issue of Concurrency and Computation: Practice and Experience, Volume 29, january 2017

8.1.3.2. Reviewer - Reviewing Activities

- Marc Shapiro, reviewer IEEE Transactions on Software Engineering.
- Swan Dubois, Theoretical Computer Science (TCS), Theory of Computing Systems (TOCS), International Journal of Networking and Computing (IJNC).
- Luciana Arantes, reviewer Journal of Parallel and Distributed Systems (JPDC).

8.1.4. Invited Talks

Pierre Sens, *Failure detection in large and dynamic distributed systems*. Univ. Curitiba, Brazil. October 2017.

Pierre Sens, *Failure detection in large and dynamic distributed systems*. Keynote speaker, GDR RSD and ASF Winter School on Distributed Systems, Pleynet, Sept Laux, Mars, 2017

Marc Shapiro, *The Antidote Cloud Database*. Comité de pilotage Groupe de Travail Logiciels Libres, Paris, Jan. 2017.

Marc Shapiro, *Just-Right Consistency*. Invited talk, joint session of PaPoC and LADIS workshops, EuroSys April 2017. Belgrade, Serbia.

Marc Shapiro, *AntidoteDB: a developer-friendly, open-source cloud database*. Invited talk, Datageeks Paris, Vente Privée la Plaine-Saint-Denis, May 2017.

Marc Shapiro, *AntidoteDB: a developer-friendly, open-source cloud database*. Invited talk, Open Source Innovation Spring, Paris May 2017.

Marc Shapiro, *Just-Right Consistency*. *La demie-heure de science*, Inria Paris, Oct. 2017.

Marc Shapiro, *AntidoteDB : Une base de données nuage pour la juste cohérence.* Paris Open Source Summit, Saint-Denis, Nov. 2017.

Marc Shapiro, *Semantics and proof of geo-replicated file system*. Invited talk, ENS Ulm, Nov. 2017.

Marc Shapiro, Invited speaker, workshop of Verification of Distributed Systems, Essaouira, May 2018.

8.1.5. Scientific Expertise

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

Marc Shapiro, reviewer for ERC Starting Grant panel PE-6.

Marc Shapiro, invited for CACM Technical Perspective [18].

Marc Shapiro, invited for several entries in Springer Encyclopedia of Database Systems [36], [37], [38] and Encyclopedia of Big Data Technologies (to appear).

Marc Shapiro, reviewer for Irish Research Council.

Marc shapiro, reviewer for Indo French Centre for the Promotion of Advanced Research (CEFIPRA).

8.1.6. Research Administration

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.

Pierre Sens, since 2015, officer at scientific research vice presidency UPMC

Pierre Sens, since 2014: Member of Steering Committee of International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD).

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

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

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

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

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

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

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

Master: Luciana Arantes, System distributed Programming, M1, UPMC Sorbonne Universités, France

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

Master: Jonathan Lejeune, Julien Sopena, Client-server distributed systems, M1, UPMC Sorbonne Universités, France

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

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

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

Licence: Swan Dubois, Franck Petit, Advanced C Programming, L2, UPMC Sorbonne Universités, France

Licence: Swan Dubois, Sébastien Monnet, Introduction to operating systems, L2, UPMC Sorbonne Universités, France

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

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

8.2.2. Supervision

PhD: Antoine Blin, “Execution of real-time applications on a small multicore embedded system”, 30 January 2017, Gilles Muller (Whisper) and Julien Sopena, CIFRE Renault

PhD: Tao Thanh Vinh, “Ensuring Availability and Managing Consistency in Geo-Replicated File Systems”, UPMC, CIFRE, 8 December 2017, Marc Shapiro, Vianney Rancurel (Scality).

PhD: Rudyar Cortes,“Un Environnement à grande échelle pour le traitement de flots massifs de données,” UPMC, funded by Chile government, 6 April 2017, Olivier Marin, Luciana Arantes, 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”, UPMC, since Oct. 2016, Swan Dubois, Franck Petit, Yoann Dieudonné (University of Picardy Jules Verne)

PhD in progress: Marjorie Bournat, “Exploration with robots in dynamic networks”, UPMC, 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.”, UPMC, CIFRE, since Jan. 2015, Sébastien Monnet, Pierre Sens, Julien Sopena, Dimitri Refauvelet (Magency).

PhD in Progress: Florent Coriat, “Géolocalisation et routage en situation de crise” since Sept 2014, UPMC, Anne Fladenmuller (NPA-LIP6) and Luciana Arantes.

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: Lyes Hamidouche, “Data replication and data sharing in mobile networks”, UPMC, CIFRE, since Nov. 2014, Sébastien Monnet, Pierre Sens, Dimitri Refauvelet (Magency).

PhD in progress: Denis Jeanneau,“Problèmes d’accord et détecteurs de défaillances dans les réseaux dynamique,” UPMC, funded by Labex Smart, since Oct. 2015, Luciana Arantes, Pierre Sens.

PhD in progress: Francis Laniel, UPMC, 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: Ilyas Toumlilt, UPMC, since Sept. 2017, advised by Marc Shapiro. “Bridging the CAP gap all the way to the edge.”

PhD in progress: Alejandro Z. Tomsic, UPMC, since Feb. 2014, Marc Shapiro. “Computing over widely-replicated data in a hybrid cloud.”

CIFRE PhD in progress: Dimitrios Vasilas, UPMC, “Indexing in large-scale storage systems.” Since Sept. 2016, advised by Marc Shapiro.

PhD in progress: Gauthier Voron, “Big-Os : un OS pour les grands volumes de données,”, UPMC, since Sep. 2014, Gaël Thomas, Pierre Sens.

8.2.3. Juries

Pierre Sens was the reviewer of:

- Jalil Boukhobza, HDR, UBO, Brest
- Adrien Lebre, HDR, EMN, Nantes
- Cédric Tedeschi, HDR, IRISA, Rennes
- Ismael Cuadrado-Cordero, PhD, IRSIA, Rennes
- Hana Teyeb, PhD, Telecom SudParis, Evry

Pierre Sens was Chair of

- Georgios Bouloukakis, PhD, UPMC-Inria, Paris, (Advisor: V. Issarny)
- José Manuel Rubio-Hernan, PhD, Telecom SudParis, Evry (Advisor: J. Garcia-Alfaro)
- Luis Eduardo Pineda Morales, PhD, Irisa, Rennes (Advisors: F. Desprez, A. Lebre)

Marc Shapiro was a member of the PhD defense committee of Paolo Viotti, EURECOM, April 2017.

8.3. Popularization

Jonathan Lejeune and Julien Sopena animated an activity during the Science Festival 2017 at UPMC

9. Bibliography

Major publications by the team in recent years

[1] V. BALEGAS, S. DUARTE, C. FERREIRA, R. RODRIGUES, N. PREGUIÇA, M. NAJAFZADEH, M. SHAPIRO. *Putting Consistency back into Eventual Consistency*, in "Euro. Conf. on Comp. Sys. (EuroSys)", Bordeaux, France, ACM, 2015, pp. 6:1–6:16 [*DOI :* 10.1145/2741948.2741972], https://hal.inria.fr/hal-01248191

[2] G. BOSILCA, A. BOUTEILLER, A. GUERMOUCHE, T. HÉRAULT, Y. ROBERT, P. SENS, J. DONGARRA. *Failure Detection and Propagation in HPC systems*, in "SC 2016 - The International Conference for High Performance Computing, Networking, Storage and Analysis", Salt Lake City, United States, November 2016, https://hal.inria.fr/hal-01352109

[3] S. DUBOIS, R. GUERRAOUI, P. KUZNETSOV, F. PETIT, P. SENS. *The Weakest Failure Detector for Eventual Consistency*, in "34th Annual ACM Symposium on Principles of Distributed Computing (PODC-2015), Donostia-San Sebastián, Spain", Donostia-San Sebastián, Spain, July 2015, pp. 375-384 [*DOI :* 10.1145/2767386.2767404], https://hal.archives-ouvertes.fr/hal-01213330

[4] L. GIDRA, G. THOMAS, J. SOPENA, M. SHAPIRO, N. NGUYEN. *NumaGiC: a Garbage Collector for Big Data on Big NUMA Machines*, in "20th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)", Istanbul, Turkey, Architectural Support for Programming Languages and Operating Systems (ASPLOS), ACM, March 2015, pp. 661-673 [*DOI :* 10.1145/2694344.2694361], https://hal.archives-ouvertes.fr/hal-01178790

[5] A. GOTSMAN, H. YANG, C. FERREIRA, M. NAJAFZADEH, M. SHAPIRO. *'Cause I'm Strong Enough: Reasoning about Consistency Choices in Distributed Systems*, in "Symposium on Principles of Programming Languages", Saint Petersburg, FL, United States, January 2016, pp. 371–384 [*DOI :* 10.1145/2837614.2837625], https://hal.inria.fr/hal-01243192

[6] M. SAEIDA ARDEKANI, P. SUTRA, M. SHAPIRO. *Non-Monotonic Snapshot Isolation: scalable and strong consistency for geo-replicated transactional systems*, in "Symp. on Reliable Dist. Sys. (SRDS)", Braga, Portugal, IEEE Comp. Society, Oct. 2013, pp. 163–172 [*DOI :* 10.1109/SRDS.2013.25], http://lip6.fr/Marc.Shapiro/papers/NMSI-SRDS-2013.pdf

[7] M. SAEIDA ARDEKANI, P. SUTRA, M. SHAPIRO. *G-DUR: A Middleware for Assembling, Analyzing, and Improving Transactional Protocols*, in "Middleware", Bordeaux, France, IEEE, December 2014, 12 p. [*DOI :* 10.1145/2663165.2663336], https://hal.inria.fr/hal-01109114

[8] Y. SAITO, M. SHAPIRO. *Optimistic Replication*, in "ACM Computing Surveys", March 2005, vol. 37, n^{o} 1, pp. 42–81, http://lip6.fr/Marc.Shapiro/papers/Optimistic_Replication_Computing_Surveys_2005-03_cameraready.pdf

[9] M. SHAPIRO, N. PREGUIÇA, C. BAQUERO, M. ZAWIRSKI. *Conflict-free Replicated Data Types*, in "Int. Symp. on Stabilization, Safety, and Security of Distributed Systems (SSS)", Grenoble, France, X. DÉFAGO, F. PETIT, V. VILLAIN (editors), Lecture Notes in Comp. Sc., Springer-Verlag, Oct. 2011, vol. 6976, pp. 386–400

[10] V. VAFEIADIS, M. HERLIHY, T. HOARE, M. SHAPIRO. *Proving Correctness of Highly-Concurrent Linearisable Objects*, in "Symp. on Principles and Practice of Parallel Prog. (PPoPP)", New York, USA, March 2006, pp. 129–136

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. BLIN. *Towards an efficient use of multi-core processors in mixed criticality embedded systems*, Université Pierre et Marie Curie - Paris VI, January 2017, https://tel.archives-ouvertes.fr/tel-01624259

[12] R. CORTÉS. *Scalable Location-Temporal Range Query Processing for Structured Peer-to-Peer Networks*, Pierre et Marie Curie, Paris VI ; LIP6 UMR 7606 UPMC Sorbonne Universités, France ; équipe REGAL, April 2017, https://tel.archives-ouvertes.fr/tel-01552377

[13] V. TAO THANH. *Ensuring Availability and Managing Consistency in Geo-Replicated File Systems*, Pierre et Marie Curie, Paris VI ; Inria Paris ; REGAL ; Scality, December 2017, https://hal.inria.fr/tel-01673030

Articles in International Peer-Reviewed Journals

[14] K. ALTISEN, A. COURNIER, S. DEVISMES, A. DURAND, F. PETIT. *Self-Stabilizing Leader Election in Polynomial Steps*, in "Information and Computation", 2017 [*DOI :* 10.1016/J.IC.2016.09.002], http://hal.upmc.fr/hal-01347471

[15] X. BONNAIRE, R. CORTÉS, F. KORDON, O. MARIN. *ASCENT: a Provably-Terminating Decentralized Logging Service*, in "The Computer Journal", December 2017, vol. 60, n° 12, pp. 1889–1911, forthcoming [*DOI :* 10.1093/COMJNL/BXX076], http://hal.upmc.fr/hal-01547514

[16] G. BOSILCA, A. BOUTEILLER, A. GUERMOUCHE, T. HÉRAULT, Y. ROBERT, P. SENS, J. DONGARRA. *A Failure Detector for HPC Platforms*, in "International Journal of High Performance Computing Applications", 2017, https://hal.inria.fr/hal-01531522

[17] T.-M.-T. NGUYEN, L. HAMIDOUCHE, F. MATHIEU, S. MONNET, S. ISKOUNEN. *SDN-based Wi-Fi Direct Clustering for Cloud Access in Campus Networks*, in "Annals of Telecommunications, Springer", 2017 [*DOI :* 10.1007/S12243-017-0598-Z], http://hal.upmc.fr/hal-01567735

[18] M. SHAPIRO. *Technical Perspective: Unexpected Connections*, in "Communications- ACM", July 2017, vol. 60, n° 8, pp. 82–82 [*DOI :* 10.1145/3068768], https://hal.inria.fr/hal-01570845

[19] P. VIOTTI, D. DOBRE, M. VUKOLIĆ. *Hybris: Robust Hybrid Cloud Storage*, in "Transactions on Storage", September 2017, vol. 13, n° 3, pp. 1 - 32 [*DOI :* 10.1145/3119896], https://hal.inria.fr/hal-01610463

International Conferences with Proceedings

[20] L. ARANTES, R. FRIEDMAN, O. MARIN, P. SENS. *Probabilistic Byzantine Tolerance Scheduling in Hybrid Cloud Environments*, in "18th International Conference on Distributed Computing and Networking (ICDCN 2017)", Hyderabad, India, January 2017 [*DOI :* 10.1145/1235], https://hal.inria.fr/hal-01399026

[21] M. BOURNAT, S. DUBOIS, F. PETIT. *Computability of Perpetual Exploration in Highly Dynamic Rings*, in "The 37th IEEE International Conference on Distributed Computing Systems (ICDCS 2017)", Atlanta, United States, June 2017, https://hal.inria.fr/hal-01548109

[22] M. BOURNAT, S. DUBOIS, F. PETIT. *Quel est le nombre optimal de robots pour explorer un anneau hautement dynamique ?* , in "ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01516182

[23] D. CARVER, J. SOPENA, S. MONNET. *ACDC : Advanced Consolidation for Dynamic Containers*, in "NCA", Cambridge, MA, United States, October 2017, https://hal.inria.fr/hal-01673304

[24] L. HAMIDOUCHE, S. MONNET, F. BARDOLLE, P. SENS, D. REFAUVELET. *EDWiN : leveraging device-to-device communications for Efficient data Dissemination over Wi-Fi Networks*, in "The 31st IEEE International Conference on. Advanced Information Networking and Applications (AINA-2017)", Taipei, Taiwan, March 2017, https://hal.inria.fr/hal-01515372

[25] L. HAMIDOUCHE, S. MONNET, P. SENS, D. REFAUVELET. *Toward heterogeneity-aware device-to-device data dissemination over Wi-Fi networks*, in "ICPADS 2017 - International Conference on Parallel and Distributed Systems", Shenzhen, China, December 2017, http://hal.univ-smb.fr/hal-01619216

[26] D. JEANNEAU, T. RIEUTORD, L. ARANTES, P. SENS. *Détecteur de fautes pour le k-accord dans les systèmes inconnus et dynamiques*, in "ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01511559

[27] J. KNIESS, L. ARANTES, P. SENS, C. V. N. ALBUQUERQUE. *Saving Resources in Discovery Protocol on Delay-Sensitive Rescue Mobile Networks*, in "The 31st IEEE International Conference on. Advanced Information Networking and Applications (AINA-2017)", Tapei, Taiwan, March 2017, https://hal.inria.fr/hal-01515369

[28] L. LE FRIOUX, S. BAARIR, J. SOPENA, F. KORDON. *PaInleSS: a Framework for Parallel SAT Solving*, in "The 20th International Conference on Theory and Applications of Satisfiability Testing", Melbourne, Australia, Lecture Notes in Computer Science, Springer, August 2017, vol. 10491, pp. 233-250 [*DOI :* 10.1007/978-3-319-66263-3_15], https://hal.archives-ouvertes.fr/hal-01540785

[29] J. LEJEUNE, F. ALVARES, T. LEDOUX. *Towards a generic autonomic model to manage Cloud Services*, in "The 7th International Conference on Cloud Computing and Services Science (CLOSER 2017)", Porto, Portugal, April 2017, https://hal.archives-ouvertes.fr/hal-01511360

[30] B. LEPERS, W. ZWAENEPOEL, J.-P. LOZI, N. PALIX, R. GOUICEM, J. SOPENA, J. LAWALL, G. MULLER. *Towards Proving Optimistic Multicore Schedulers*, in "HotOS 2017 - 16th Workshop on Hot Topics in Operating Systems", Whistler, British Columbia, Canada, ACM SIGOPS, May 2017, 6 p. [*DOI :* 10.1145/3102980.3102984], https://hal.inria.fr/hal-01556597

[31] B. NGOM, M. MAKPANGOU. *Summary Prefix Tree: An over DHT Indexing Data Structure for Efficient Superset Search*, in "NCA 2017 : 16th IEEE International Symposium on Network Computing and Applications", Cambridge, MA, United States, October 2017, https://hal.inria.fr/hal-01672052

[32] J. PAULO DE ARAUJO, L. ARANTES, E. P. DUARTE, L. A. RODRIGUES, P. SENS. *A Publish/Subscribe System Using Causal Broadcast Over Dynamically Built Spanning Trees*, in "SBAC-PAD 2017 - 29th International Symposium on Computer Architecture and High Performance Computing", Campinas, Brazil, IEEE, October 2017, pp. 161-168 [*DOI :* 10.1109/SBAC-PAD.2017.28], https://hal.inria.fr/hal-01644469

[33] W. SUN, V. SIMON, S. MONNET, P. ROBERT, P. SENS. *Analysis of a Stochastic Model of Replication in Large Distributed Storage Systems: A Mean-Field Approach*, in "ACM Sigmetrics 2017- International Conference on Measurement and Modeling of Computer Systems", Urbana-Champaign, Illinois, United States, ACM, June 2017, pp. 51–51, https://arxiv.org/abs/1701.00335 [*DOI :* 10.1145/3078505.3078531], https://hal.inria.fr/hal-01494235

[34] G. VORON, G. THOMAS, V. QUEMA, P. SENS. *An interface to implement NUMA policies in the Xen hypervisor*, in "Twelfth European Conference on Computer Systems, EuroSys 2017", Belgrade, Serbia, April 2017, 15 p. , https://hal.inria.fr/hal-01515359

Conferences without Proceedings

[35] S. BOUCHARD, M. BOURNAT, Y. DIEUDONNÉ, S. DUBOIS, F. PETIT. *Asynchronous Approach in the Plane: A Deterministic Polynomial Algorithm*, in "31st International Symposium on Distributed Computing, DISC 2017", Vienna, Austria, October 2017, http://hal.upmc.fr/hal-01672916

Scientific Books (or Scientific Book chapters)

[36] M. SHAPIRO, B. KEMME. *Eventual Consistency*, in "Encyclopedia of Database Systems", L. LIU, M. T. ÖZSU (editors), Springer, June 2017, 2 p. [*DOI :* 10.1007/978-1-4899-7993-3_1366-2], https://hal.inria.fr/hal-01547451

[37] M. SHAPIRO. *Optimistic Replication and Resolution*, in "Encyclopedia Of Database Systems", L. LIU, M. T. ÖZSU (editors), Springer-Verlag, April 2017, vol. Optimistic Replication and Resolution, pp. 1–8 [*DOI :* 10.1007/978-1-4899-7993-3_258-4], https://hal.inria.fr/hal-01576333

[38] M. SHAPIRO. *Replicated Data Types*, in "Encyclopedia Of Database Systems", L. LIU, M. T. ÖZSU (editors), Springer-Verlag, July 2017, vol. Replicated Data Types, pp. 1–5 [*DOI :* 10.1007/978-1-4899-7993-3_80813-1], https://hal.archives-ouvertes.fr/hal-01578910

Research Reports

[39] G. BOSILCA, A. BOUTEILLER, A. GUERMOUCHE, T. HÉRAULT, Y. ROBERT, P. SENS, J. DONGARRA. *A Failure Detector for HPC Platforms*, Inria, February 2017, n° RR-9024, https://hal.inria.fr/hal-01453086

[40] E. MAUFFRET, D. JEANNEAU, L. ARANTES, P. SENS. *The Weakest Failure Detector to Solve the Fault Tolerant Mutual Exclusion Problem in an Unknown Dynamic Environment*, LISTIC ; Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, December 2017, https://hal.archives-ouvertes.fr/hal-01661127

[41] S. S. NAIR. *Evaluation of the CEC (Correct Eventual Consistency) Tool*, Inria Paris ; LIP6 UMR 7606, UPMC Sorbonne Universités, France, November 2017, n° RR-9111, pp. 1-27, https://hal.inria.fr/hal-01628719

Other Publications

[42] P. DARCHE. *Évolution des mémoires à semi-conducteurs à accès aléatoire*, February 2017, Article E2491 des Techniques de l'Ingénieur-article de référence sur le domaine, http://hal.upmc.fr/hal-01341972

[43] A. SINGH, M. SHAPIRO, G. THOMAS. *Persistent Memory Programming Abstractions in Context of Concurrent Applications*, December 2017, https://arxiv.org/abs/1712.04989 - Accepted in HiPC SRS 2017, https://hal.inria.fr/hal-01667772