IN PARTNERSHIP WITH:
CNRS

Université Pierre et Marie Curie (Paris 6)

Activity Report 2015

Project-Team REGAL

Large-Scale Distributed Systems and Applications

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

RESEARCH CENTER
Paris - Rocquencourt

THEME
Distributed Systems and middleware

Table of contents

Project-Team REGAL

Creation of the Project-Team: 2005 July 01

Keywords:

Computer Science and Digital Science:

1.1.1. - Multicore
1.1.13. - Virtualization
1.1.6. - Cloud
1.1.7. - Peer to peer
1.1.9. - Fault tolerant systems
1.3. - Distributed Systems
1.6. - Green Computing
2.6. - Infrastructure software
2.6.1. - Operating systems
2.6.2. - Middleware
2.6.3. - Virtual machines
3.1.3. - Distributed data
3.1.8. - Big data (production, storage, transfer)
7.1. - Parallel and distributed algorithms

Other Research Topics and Application Domains:

4.4. - Energy consumption
6.4. - Internet of things
8.2. - Connected city
9.2.3. - Video games
9.4.1. - Computer science

1. Members

Research Scientists

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

Faculty Members

Pierre Sens [Team leader, UPMC, Professor, HdR]
Luciana Bezerra Arantes [UPMC, Associate Professor]
Swan Dubois [UPMC, Associate Professor]
Olivier Marin [UPMC, Associate Professor, until July 2015]
Sébastien Monnet [UPMC, Associate Professor, HdR]
Franck Petit [UPMC, Professor, HdR]
Julien Sopena [UPMC, Associate Professor]

Engineers

Tyler Crain [Inria, granted by FP7 -SYNCFREE- project]
Salvatore Pileggi [Inria]
Marek Zawirski [Inria, until Mar 2015, granted by Google Inc]

PhD Students

Marjorie Bournat [UPMC]
Damien Carver [UPMC, CIFRE with Magency]
Rudyar Cortes [UPMC]
Lokesh Gidra [Research Engineer at Hewlett Packard Enterprise, until Sep 2015]
Lyes Hamidouche [UPMC, CIFRE with Magency]
Mohamed Hamza Kaaouachi [UPMC, until Sep 2015]
Denis Jeanneau [UPMC]
Maxime Lorrillere [UPMC]
Rostom Mennour [Univ. de Constantine 2, Doctoral Internship, from Oct 2015]
Mahsa Najafzadeh [Inria]
Bassirou Ngom [U. Cheikh Anta Diop Dakar, Sénégal, Joint PhD Student,]
Karine Pires [UPMC, Telecom Bretagne until Mar. 2015]
Vinh Tao Thanh [UPMC, CIFRE with Scality]
Alejandro Tomsic [Inria, granted by FP7 -SYNCFREE- project]
Guillaume Turchini [UPMC]
Maxime Véron [CNAM, until Sep. 2015]
Gauthier Voron [UPMC]
Marek Zawirski [Inria, until Jan. 2015, granted by FP7 -SYNCFREE- project]

Administrative Assistant

Hélène Milome [Inria]

Others

Santiago Javier Alvarez Colombo [Inria, Masters' Internship, from Jul 2015]
Thibault Rieutord [Normale Sup Rennes, Internship, until Jul 2015]

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 and Inria Paris-Rocquencourt. In 2014, 4 permanent members of Regal created Whisper team with a focuss on infrastructure (system) software.

3. Research Program

3.1. Research rationale

As society relies more and more on computers, responsiveness, correctness and security are increasingly critical. At the same time, systems are growing larger, more parallel, and more unpredictable. Our research agenda is to design Computer Systems that remain correct and efficient despite this increased complexity and in spite of conflicting requirements. The term *"Computer Systems"* is interpreted broadly,[1] and includes system architecture, operating systems, distributed systems, multiprocessor systems, and touches on related areas such as computer networks, distributed databases or support for big data. The interests of the Regal group cover the whole spectrum from theory to experimentation, with a strong focus on algorithm design and implementation.

[1] This follows the definition from the journal of reference in our field, ACM Transactions on Computer Systems.

This holistic approach allows us to address related problems at different levels. It also permits us to efficiently share knowledge and expertise, and is a source of originality.

Computer Systems is a rapidly evolving domain, with strong interactions with industry. Two main evolutions in the Computer Systems area have strongly influenced our research activities:

3.1.1. Modern computer systems are increasingly parallel and distributed.

Ensuring the 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. Ease of use, performance and efficiency are equally important for systems to be accepted. These requirements are somewhat conflicting, and there are many algorithmic and engineering trade-offs, which often depend on specific workloads or usage scenarios.

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 system state can only be partial and incomplete. The challenges of distributed computing listed above are subject to new trade-offs.

Innovative environments that motivate our research include cloud computing, geo-replication, edge clouds, peer-to-peer (P2P) systems, dynamic networks, and manycore machines. The scientific challenges are scalability, fault tolerance, security, dynamicity and the virtualization of the physical infrastructure. Algorithms designed for classical distributed systems, such as resource allocation, data storage and placement, and concurrent and consistent access to shared data, need to be revisited to work properly under the constraints of these new environments.

Regal focuses in particular on two key challenges in these areas: the adaptation of algorithms to the new dynamics of distributed systems and data management on large configurations.

3.1.2. Multicore architectures are everywhere.

The fine-grained parallelism offered by multicore architectures has the potential to open highly parallel computing to new application areas. To make this a reality, however, many issues, including issues that have previously arisen in distributed systems, need to be addressed. Challenges include obtaining a consistent view of shared resources, such as memory, and optimally distributing computations among heterogeneous architectures, such as CPUs, GPUs, and other specialized processors. As compared to distributed systems, in the case of multicore architectures, these issues arise at a more fine-grained level, leading to the need for different solutions and different cost-benefit trade-offs.

Of particular interest to Regal are topics related to memory management in high-end multicore computers, such as garbage collection of very large memories and system support for massive databases of highly-structured data.

4. Highlights of the Year

4.1. Highlights of the Year

- *Garbage collection for big data on large-memory NUMA machines.* We developed NumaGiC, a high-throughput garbage collector for big-data algorithms running on large-memory NUMA machines. This result, a collaboration with the Whisper team, has been presented at ASPLOS 2015 [49].
- *Explicit consistency.* We propose an alternative approach to the strong-vs.-weak consistency conundrum, *explicit consistency*. This result has been presented at EuroSys 2015 [80]. We have also developed a new sound logic for proving the correctness of a distributed database under concurrent updates. This result is published at POPL 2016 [50].

- *The weakest failure detector of implement eventual consistency*. We found the weakest failure detector to implement an eventually consistent replicated service. This theoretical result has been presented at PODC 2015 [46].

4.1.1. Awards

Gauthier Voron obtained best paper award at system track of Compas'2015.

BEST PAPER AWARD:

[64]

V. GAUTHIER, G. THOMAS, P. SENS, V. QUEMA. *Optimisation mémoire dans une architecture NUMA : comparaison des gains entre natif et virtualisé*, in "Conférence en Parallélisme, Architecture et Système, (COMPAS'15)", Lille, France, 2015, https://hal.inria.fr/hal-01253189

5. New Software and Platforms

5.1. Antidote

FUNCTIONAL DESCRIPTION

Antidote is the flexible cloud database platform currently under development in the SyncFree European project. 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 [37], [50], for elastic version management, for adaptive replication, for partial replication, and for reconfigurable sharding.

- Participants: Tyler Crain, Marc Shapiro, Serdar Tasiran and Alejandro Tomsic
- Contact: Tyler Crain
- URL: https://github.com/SyncFree

5.2. G-DUR

FUNCTIONAL DESCRIPTION

A large family of distributed transactional protocols have a common structure, called Deferred Update Replication (DUR). DUR provides dependability by replicating data, and performance by not re-executing transactions but only applying their updates. Protocols of the DUR family differ only in behaviors of few generic functions. Based on this insight, we offer a generic DUR middleware, called G-DUR, along with a library of finely-optimized plug-in implementations of the required behaviors.

- Participants: Marc Shapiro, Alejandro Tomsic
- Contact: Marc Shapiro
- URL: https://github.com/msaeida/jessy

5.3. NumaGIC

FUNCTIONAL DESCRIPTION

NumaGiC is a version of the HotSpot garbage collector (GC) adapted to many-core computers with very large main memories. In order to maximise GC throughput, it manages the trade-off between memory locality (local scans) and parallelism (work stealing) in a self-balancing manner. Furthemore, the collector features several memory placement heuristics that improve locality.

- Participants: Lokesh Gidra, Marc Shapiro, Julien Sopena and Gaël Thomas
- Contact: Marc Shapiro
- URL: https://scm.gforge.inria.fr/anonscm/git/transgc/.

5.4. SwiftCloud

Functional Description

Client-side (e.g., mobile or in-browser) apps need local access to shared cloud data, but current technologies either do not provide fault-tolerant consistency guarantees, or do not scale to high numbers of unreliable and resource-poor clients, or both. Addressing this issue, the SwiftCloud distributed object database supports high numbers of client-side partial replicas. SwiftCloud offers fast reads and writes from a causally-consistent client-side cache. It is scalable, thanks to small and bounded metadata, and available, tolerating faults and intermittent connectivity by switching between data centres. The price to pay is a modest amount of staleness. A recent Inria Research Report (submitted for publication) presents the SwiftCloud algorithms, design, and experimental evaluation, which shows that client-side apps enjoy the same guarantees as a cloud data store, at a small cost.

- Participants: Marc Shapiro, Serdar Tasiran, Marek Zawirski and Mahsa Najafzadeh
- Contact: Marc Shapiro
- URL: git+ssh://scm.gforge.inria.fr//gitroot/swiftcloud

5.5. PUMA

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, Sébastien Monnet, Pierre Sens, Julien Sopena
- Contact: Maxime Lorrillere
- URL: https://github.com/mlorrillere/puma

6. New Results

6.1. Distributed algorithms for dynamic networks

Participants: Luciana Bezerra Arantes [correspondent], Marjorie Bournat, Swan Dubois, Denis Jeanneau, Mohamed Hamza Kaaouachi, Sébastien Monnet, 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.

We have obtained results both on fundamental aspects of distributed algorithms and on specific emerging large-scale applications.

We study various key topics of distributed algorithms: agreement, failure detection, data dissemination and data finding in large scale systems, self-stabilization and self-* services.

6.1.1. Agreement and failure detection in dynamic Distributed Systems

Distributed systems should provide reliable and continuous services despite the failures of some of their components. A classical way for a distributed system to tolerate failures is to detect them and then to recover. It is now well recognized that the dominant factor in system unavailability lies in the failure detection phase. In 2015, we obtain the following results on failure detection:

Assuming a message-passing environment with a majority of correct processes, the necessary and sufficient information about failures for implementing a general state machine replication scheme ensuring consistency is captured by the Ω failure detector. We show in [46] that in such a message-passing environment, Ω is also the weakest failure detector to implement an eventually consistent replicated service, where replicas are expected to agree on the evolution of the service state only after some (a priori unknown) time.

We also study the k-set agreement problem is a generalization of the consensus problem where processes can decide up to k different values. Very few papers have tackled this problem in dynamic networks. Exploiting the formalism of the Time Varying Graph model, we propose in [70]a new quorum-based failure detector for solving k-set agreement in dynamic networks with asynchronous communications. We present two algorithms that implement this new failure detector using graph connectivity and message pattern assumptions. We also provide an algorithm for solving k-set agreement using our new failure detector.

We propose several algorithms to implement efficient failure detection services. We introduce in [60] the Two Windows Failure Detector (2WFD), an algorithm that provides QoS and is able to react to sudden changes in network conditions, a property that currently existing algorithms do not satisfy. We ran tests on real traces and compared the 2W-FD to state-of-the-art algorithms. Our results show that our algorithm presents the best performance in terms of speed and accuracy in unstable scenarios. In [62], we propose a new approach towards the implementation of failure detectors for large and dynamic networks: we study reputation systems as a means to detect failures. The reputation mechanism allows efficient node cooperation via the sharing of views about other nodes. Our experimental results show that a simple prototype of a reputation-based detection service performs better than other known adaptive failure detectors, with improved flexibility. It can thus be used in a dynamic environment with a large and variable number of nodes.

6.1.2. Probabilistic Byzantine Tolerance allocation strategies in Hybrid Cloud Environments

We explore the node allocation challenges in providing probabilistic Byzantine fault tolerance in a hybrid cloud environment, consisting of nodes with varying reliability levels, compute power, and monetary cost. We consider hybrid computing architectures that combine edge nodes with cloud hosted computing. In such a system, a large fraction of the computation is performed by donated machines at the edge of the network, which significantly reduces the cost to the owner of the computation.

Considering “bag of tasks” (BoT) applications where a large computational problem is broken into a large number of independent tasks, the probabilistic Byzantine fault tolerance guarantee refers to the confidence level that the result of a given computation is correct despite potential Byzantine failures. In [36] we explore probabilistic Byzantine tolerance, in which computation tasks are replicated on dynamic replication sets whose size is determined based on ensuring probabilistic thresholds of correctness.

6.1.3. Covering problems in dynamic systems

We study covering problems (such as minimal dominating set or maximal matching) in the context of highly dynamic distributed systems. We first obtain some general results. In [48], we first propose a new definition of this family of problems since classical ones are meaningless in such systems. We generalize the classical definition of time complexity (for static systems) to our setting. We also provided in [40] a generic tool to help the writing of impossibility proofs in dynamic distributed systems. Then, we focus on the particular case of the minimal dominating set problem. We characterize the necessary and sufficient condition to construct deterministically a minimal dominating set in a dynamic system according to our definition.

6.1.4. Self-Stabilization

Self-stabilization is a generic paradigm to tolerate transient faults (*i.e.*, faults of finite duration) in distributed systems. Results obtained in this area by Regal members in 2015 follow.

Spanning tree construction is a well-studied problem in distributed computing for its numerous applications like routing, broadcast...Properties of the obtained trees, efficiency of the construction, and fault-tolerance guarantees are naturally at the heart of many researches. In this context, we propose in [39] a new self-stabilizing algorithm for the minimum diameter spanning tree that achieves better time and space complexity than existing solutions. Moreover, our solution tolerates a fully asynchronous adversary.

A classical way to endowed self-stabilization with (permanent) fault tolerance is *confinement*. That is, we ensure that the self-stabilizing system moreover ensures that the effect of permanent faults is limited to some topological areas of the system. In [27], we propose a characterization of optimal confinement areas for a large set of spanning tree metrics in presence of Byzantine faults. In [24], we propose a stabilizing implementation of an atomic register in presence of crash faults. By avoiding the propagation of fault effects further than a given radius, confinement is clearly a *spatial* approach. Another approach, called *temporal*, consists in recovering as quick as possible to a configuration from which some forms of safety are satisfied.

In [68], we introduce the notion of *gradual stabilization* and provide a gradually self-stabilizing algorithm that solves the *unison* problem, *i.e.*, the problem that consists in synchronizing logical clocks locally maintained by the processes.

6.1.5. Team of Mobile Robots

Swarm of autonomous mobile sensor devices (or, robots) recently emerged as an attractive issue in the study of dynamic distributed systems permits to assess the intrinsic difficulties of many fundamentals tasks, such as exploring or gathering in a discrete space. We consider autonomous robots that are endowed with visibility sensors (but that are otherwise unable to communicate) and motion actuators. The robots we consider are weak, *i.e.*, they are anonymous, uniform, unable to explicitly communicate, and oblivious (they do not remember any of their past actions). Despite their weakness, those robots must collaborate to solve a collective tasks such as exploration, gathering, flocking, to quote only a few.

In [45], we first show that it is impossible to explore any simple torus of arbitrary size with (strictly) less than four robots, even if the algorithm is probabilistic. Next, we propose an optimal (*w.r.t.* the number of robots) solution for the terminating exploration of torus-shaped networks by a team of k such robots in the SSYNC model. The proposed algorithm is probabilistic and works for any simple torus of size $\ell \times L$, where $7 \leq \ell \leq L$. Since the optimal number of robots is also four in rings, our result shows that increasing the number of possible symmetries in the network (due to increasing dimensions) does not necessarily come at an extra cost *w.r.t.* the number of robots that are necessary to solve the problem.

6.2. Management of distributed data

Participants: Rudyar Cortes, Mesaac Makpangou, Olivier Marin, Sébastien Monnet [correspondent], Pierre Sens.

6.2.1. Long term durability and storage load distribution

In 2014, we had proposed SPLAD (for Scattering and PLAcing Data replicas to enhance long-term durability), a model that allows us to vary the data scattering degree by tuning a selection range width. We have enhanced our model [57] and we have focused on the study of the policy used while choosing a storing node within the selection range. Some policies may lead to heavily unbalanced storage load distribution which can be harmful for the system. Simple policies to balance the load (e.g. storing new blocks on least loaded nodes) may induce network congestion and thus data losses. We have shown that the "power of two choices" policy (choosing the least loaded node among two random ones) brings good results both in terms of storage load distribution and fault tolerance.

6.2.2. Management of dynamic big data

Managing and processing Dynamic Big Data, where multiple sources produce new data continuously, is very complex. Static cluster- or grid-based solutions are prone to induce bottleneck problems, and are therefore ill-suited in this context. Our objective in this domain is to design and implement a Reliable Large Scale Distributed Framework for the Management and Processing of Dynamic Big Data. In 2015, we focused on Spatio-temporal range queries over Big Location Data aim to extract and analyze relevant data items generated around a given location and time. They require concurrent processing of massive and dynamic data flows. We proposed a scalable architecture for continuous spatio-temporal range queries built by coalescing multiple computing nodes on top of a Distributed Hash Table. The key component of our architecture is a distributed spatio-temporal indexing structure which exhibits low insertion and low index maintenance costs. We assessed our solution with a public data set released by Yahoo! which comprises millions of geotagged multimedia files [43].

6.3. CISE Logic and tool for proving invariants in distributed databases

Participants: Marc Shapiro [correspondent], Mahsa Najafzadeh, Alexey Gotsman, Carla Ferreira.

We have developed a new sound logic for proving the correctness of a distributed database under concurrent updates, showing whether the application maintains the database's *integrity invariants*. An operation of the application is specified as a *preparator*, which checks the operation's precondition at an origin replica and generates an *effector*. The effector abstracts the update to be applied to every replica. The application also specifies which operations are allowed to take place concurrently. In summary, the logic shows that the application maintains the invariant if the three following rules are satisfied:

- Each operation individually maintains the invariant. It follows that operations' preconditions are sufficiently strong to ensure correctness in a sequential execution.
- The effectors of any two operations that can execute concurrently commute. This implies that the database replicas all converge to the same state.
- For any pair of operations u and v that can execute concurrently, the precondition of u is stable under the effector of v, and vice-versa.

This result is published at POPL 2016 [50].

We have implemented a tool (based on the Z3 SMT solver) that implements these rules. A demo of the tool is available online [78]. If the application passes the tool, it is correct. If not, the tool returns a counter-example, which the application developer can inspect to find the source of the error. Generally speaking, the developer can either weaken the invariants or the effects of operations, or strengthen consistency by disallowing concurrency. By choosing one or the other, the developer performs a co-design of the application with its consistency protocol, in order to have the highest possible concurrency that still ensures correctness.

For instance, consider a database of bank accounts, with the invariant that an account's balance must be positive. The banking application has operations $credit(acct, amt)$, $debit(acct, amt)$, and $accrue - interest(acct)$. The first rule dictates that $debit$ has the precondition $amt = balance$. The second rule dictates that $accrue - interest$ computes the amount of interest according to the state at the origin, not at every replica. The third rule is violated if concurrent $debit$s are allowed; if the bank wishes to uphold the invariant, the only correct solution is to disallow concurrent $debit$s.

6.4. Memory management for big data

Participants: Antoine Blin, Damien Carver, Maxime Lorrillere, Sébastien Monnet, Julien Sopena [correspondent].

6.4.1. Automated file cache pooling

Some applications, like online sales servers, intensively use disk I/Os. Their performance is tightly coupled with I/Os efficiency. To speed up I/Os, operating systems use free memory to offer caching mechanisms. Several I/O intensive applications may require a large cache to perform well. However, nowadays resources are virtualized. In clouds, for instance, virtual machines (VMs) offer both isolation and flexibility. This is the foundation of cloud elasticity, but it induces fragmentation of the physical resources, including memory. This fragmentation reduces the amount of available memory a VM can use for caching I/Os. Previously, we proposed Puma (for Pooling Unused Memory in Virtual Machines) which allows I/O intensive applications running on top of VMs to benefit of large caches. This was realized by providing a remote caching mechanism that provides the ability for any VM to extend its cache using the memory of other VMs located either in the same or in a different host.

We have performed an extensive evaluation of Puma [53] and we have enhanced our solution: Puma adapts automatically the amount a memory that a VM offers to another VM. Furthermore, if the network becomes overloaded, Puma detects a performance degradation and stops using a remote cache.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. Joint industrial PhD with Orange Labs and Renault

- Orange Lab, 30,000 euros for 1 PhD Students (CIFRE), Ralucca Diaconu
- Renault, 60,000 over 3 years (2013 - 2016) for a CIFRE. In the context of a Cifre cooperation with Renault, we are supervising with Whipser the PhD of Antoine Blin on the topic of scheduling processes on a multicore machine for the automotive industry. The goal is to allow real-time and multimedia applications to cohabit on a single processor. The challenge here is to control resource consumption of non real-time processes so as to preserve the real-time behavior of critical ones. As part of this cooperation, we will use the Bossa DSL framework for implementing process schedulers that we have previously developed.

7.1.2. Joint industrial PhD: CRDTs for Large-Scale Storage Systems, with Scality SA

This year, we continued the joint CIFRE (industrial PhD) research of Tao Thanh Vinh, with the French start-up company Scality, as described above (under “Large-Scale File Systems”).

The objective of this research is to design new algorithms for file and block storage systems, considering both the issues of scaling the file naming tree to a very large size, and the issue of conflicting updates to files or to the name tree, in the case of high latency or disconnected work. Preliminary results were published at Systor 2015 [58].

7.1.3. EMR CREDIT, with Thales.

Franck Petit and Swan Dubois participate to the creation of the EMR (Equipe Mixte de Recherche) *CREDIT*, (Compréhension, Représentation et Exploitation Des Interactions Temporelles) between LIP6/UPMC and Thales.

Nowadays, networks are the field of temporal interactions that occur in many settings networks, including security issues. The amount and the speed of such interactions increases everyday. Until recently, the dynamics of these objects was little studied due to the lack of appropriate tools and methods. However, it becomes crucial to understand the dynamics of these interactions. Typically, how can we detect failures or attacks in network traffic, fraud in financial transactions, bugs or attacks traces of software execution. More generally, we seek to identify patterns in the dynamics of interactions. Recently, several different approaches have been proposed to study such interactions. For instance, by merging all interactions taking place over a period (e.g. one day) in a graph that are studied thereafter (evolving graphs). Another approach was to built meta-objects by duplicating entities at each unit of time of their activity, and by connecting them together.

The goal of the EMR is to join both teams of LIP6 and Thales on these issues. More specifically, we hope to make significant progress on security issues such as anomaly detection. This requires the use of a formalism sufficiently expressive to formulate complex temporal properties. Recently, a vast collection of concepts, formalisms, and models has been unified in a framework called Time-Varying Graphs. We want to pursuit that way. In the short run, the challenges facing us are: (1) refine the model to capture some interaction patterns, (2) design of algorithms to separate sequences of interactions, (3) Identify classes of entities playing a particular role in the dynamics, such as bridges between communities, or sources and sinks.

7.1.4. Joint industrial PhDs: data sharing in mobile networks and automatic resizing of shared I/O caches, with Magency

Magency organizes large events during which participants can use mobile devices to access related data and interact together.

The thesis of Lyes Hamidouche concerns efficient data sharing among a large number of mobile devices. Magency brings traces captured during real events (data accesses and user mobility). We are jointly working on the design of algorithms allowing a large number of mobile devices to efficiently access remote data.

Magency also runs servers. A server is used before an event in order to be prepared and tested, and then, during the event to serve the numerous mobile devices accesses. Many servers are run on a single physical machine using containers. Using this configuration, the memory is partitioned, leading to poor performances for applications that need a large amount of memory for caching purpose. In the context of Damien Carver's PhD thesis, we are designing kernel-level mechanisms that automatically give more memory to the most active containers, leveraging the expertise acquired during Maxime Lorrillere's PhD thesis.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. 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 2,5 years.

8.1.2. InfraJVM - (2012–2015)

Members: LIP6 (Regal), Ecole des Mines de Nantes (Constraint), IRISA (Triskell), LaBRI (LSR).

Funding: ANR Infra.

Objectives: The design of the Java Virtual Machine (JVM) was last revised in 1999, at a time when a single program running on a uniprocessor desktop machine was the norm. Today's computing environment, however, is radically different, being characterized by many different kinds of computing devices, which are often mobile and which need to interact within the context of a single application. Supporting such applications, involving multiple mutually untrusted devices, requires resource management and scheduling strategies that were not planned for in the 1999 JVM design. The goal of InfraJVM is to design strategies that can meet the needs of such applications and that provide the good performance that is required in an MRE.

The coordinator of InfraJVM is Gaël Thomas, who left the team in 2014. Infra-JVM brings a grant of 202 000 euros from the ANR to UPMC over three years.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

8.2.1.1. SyncFree

Title: Large-scale computation without synchronisation

Programm: FP7

Duration: October 2013 - September 2016

Coordinator: Inria

Partners:

Basho Technologies (United Kingdom)

Faculdade de Ciencias e Tecnologia da Universidade Nova de Lisboa (Portugal)

Koç University (Turkey)

Rovio Entertainment Oy (Finland)

Trifork As (Denmark)

Université Catholique de Louvain (Belgium)

Technische Universitaet Kaiserslautern (Germany)

Inria contact: Marc Shapiro

The goal of SyncFree is to enable large-scale distributed applications without global synchronisation, by exploiting the recent concept of Conflict-free Replicated Data Types (CRDTs). CRDTs allow unsynchronised concurrent updates, yet ensure data consistency. This revolutionary approach maximises responsiveness and availability; it enables locating data near its users, in decentralised clouds. Global-scale applications, such as virtual wallets, advertising platforms, social networks, online games, or collaboration networks, require consistency across distributed data items. As networked users, objects, devices, and sensors proliferate, the consistency issue is increasingly acute for the software industry. Current alternatives are both unsatisfactory: either to rely on synchronisation to ensure strong consistency, or to forfeit synchronisation and consistency altogether with ad-hoc eventual consistency. The former approach does not scale beyond a single data centre and is expensive. The latter is extremely difficult to understand, and remains error-prone, even for highly-skilled programmers. SyncFree avoids both global synchronisation and the complexities of ad-hoc eventual consistency by leveraging the formal properties of CRDTs. CRDTs are designed so that unsynchronised concurrent updates do not conflict and have well-defined semantics. By combining CRDT objects from a standard library of proven datatypes (counters, sets, graphs, sequences, etc.), large-scale distributed programming is simpler and less error-prone. CRDTs are a practical and cost-effective approach. The SyncFree project will develop both theoretical and practical understanding of large-scale synchronisation-free programming based on CRDTs. Project results will be new industrial applications, new application architectures, large-scale evaluation of both, programming models and algorithms for large-scale applications, and advanced scientific understanding.

8.3. International Initiatives

8.3.1. Inria International Labs

Inria Chile
Associate Team involved in the International Lab:

8.3.1.1. ARMADA

Title: hARnessing MAssive DAta flows

International Partner (Institution - Laboratory - Researcher):

Universidad Tecnica Federico Santa Maria (Chile) - Department of Computer Science (Department of Comput) - Xavier Bonnaire

Start year: 2014

See also: http://web.inria-armada.org

The ARMADA project aims at designing and implementing a reliable framework for the management and processing of massive dynamic dataflows. The project is two-pronged: fault-tolerant middleware support for processing massive continuous input, and a redundant storage service for mutable data on a massive scale.

8.3.2. Participation In other International Programs

8.3.2.1. PHC Maimonide

Title: Application dependent intrusion (byzantine) detection in Dynamic cloud systems

International Partner (Institution - Laboratory - Researcher):

Technion Haifa - Prof. Roy Friedman

Duration: 2014–2015

The goal of this project is to study the ability to tolerate Byzantine failures in dynamic environments. The Byzantine model allows arbitrary behaviour of a certain fraction of nodes. Our goal is to provide both a theoretical framework and performance evaluation to tolerate Byzantine behaviour in dynamic distributed environments. We consider "bag of tasks" (BoT) applications characterized by trivial parallelism where a large computational problem is broken into a large number of independent tasks. These tasks can be spread on commodity hardware and operating systems. We target different executions environments: (1) Clouds: tasks are submitted to virtual machines hosted at cloud providers, (2) Desktop grid: tasks are submitted to federate large pool of donated machines hosted at user home, (3) Hybrid cloud: combining both cloud and desktop nodes.

8.3.2.2. 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.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

8.4.1.1. Internships

Dastagiri Reddy MalikiReddy

Date: May—Aug. 2015

Institution: IITKGP (India)

Alvarez Colombo Santiago Javier

Date: Jul. 2015—Jan. 2016

Institution: Universidad de Buenos Aires (Argentina)

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. General chair, scientific chair

- Luciana Arantes was scientific co-chair of the 16th Simposio em Sistemas Computacionais de Alto Desempenho
- Pierre Sens was general-chair of EDCC 2015 conference.

9.1.1.2. Member of the organizing committees

- Luciana Arantes was organisation chair of EDCC 2015 conference.
- Sébastien Monnet was finance chair for EDCC 2015 conference.
- Franck Petit is member of the steering committee of SSS (International Symposium on Stabilization, Safety, and Security of Distributed Systems) conference.
- Pierre Sens is member of the steering committee of SBAC-PAD (International Symposium on Computer Architecture and High Performance Computing) conference.
- Marc Shapiro is a member of the Steering Committee of the Principles and Practice of Consistency for Distributed Data (PaPoC).
- Marc Shapiro is a member of the Steering Committee of the Int. Conf. on the Principles of Distributed Systems (OPODIS).

9.1.2. Scientific events selection

9.1.2.1. Chair of conference program committees

- Marc Shapiro chairs the 2016 Franco-American Doctoral Exchange Programme (FADEx).

9.1.2.2. Member of the conference program committees

- Luciana Arantes was a PC member of the 15th IFIP Distributed Applications and Interoperable Systems Conference (DAIS 2015); The 26th IEEE International Symposium on Software Reliability Engineering (ISSRE 2015); the 17th IEEE International Conference on High Performance Computing and Communications (HPCC 2015); The 14th IEEE International Symposium on Network Computing and Applications (NCA 2015); The 27th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD 2015); The Eighth International Conference on Dependability (DEPEND 2015).
- Swan Dubois was member of the program committee of 17th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS 2015).
- Sébastien Monnet is PC member of 30th and 31st ACM/SIGAPP Symposium On Applied Computing - track Operating Systems (SAC 2015 and 2016); French conference on parallelism, architecture and system (Compas 2015); 14th workshop on Network and Systems Support for Games (NetGames 2015) and 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid 2016).

- Pierre Sens was member of the program committee of 4rd IEEE/SAE International Conference on Connected Vehicles and Expo (ICCVE 2015); Europar 2015; 30th IEEE International Parallel and Distributed Processing Symposium (IPDPS'2016).
- Marc Shapiro was a member of the External Programme Committee for the Int. Conf. on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2016).
- Marc Shapiro is PC member of W. on Planetary-Scale Distributed Systems (W-PSDS) 2015; member of the Middleware 2015 conference.

9.1.2.3. Reviewer

- Swan Dubois reviewed papers for the ACM Symposium on Principles of Distributed Computing (PODC 2015), the International Conference on Networked Systems (NETYS 2015), and the Rencontres Francophones sur les Aspects Algorithmiques de Télécommunications (AlgoTel 2015).
- Sébastien Monnet was external reviewer for the 30th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2016).

9.1.3. Journal

9.1.3.1. Member of the editorial boards

- Franck Petit was invited editor with Michel Raynal for a special issue on Distributed Computing into TCS (Journal of Theoretical Computer Science)—Vol. 561. Also, he belongs to the editorial board of the Scientific World Journal and the Journal of Discrete Mathematics.
- Pierre Sens is associated editor of International Journal of High Performance Computing and Networking (IJHPCN).

9.1.3.2. Reviewer - Reviewing activities

- Luciana Arantes reviewed papers for the Journal of Parallel and Distributed Computing (JPDC).
- Swan Dubois reviewed papers for the Journal of Parallel and Distributed Computing (JPDC), Computer Networks (COMNET), and Theoretical Computer Science (TCS).
- Sébastien Monnet reviewed a paper for the IEEE Transactions on Parallel & Distributed Systems (TPDS 2016).

9.1.4. Invited talks

Pierre Sens gave the following invited talks:

- ICL Innovative Computing Laboratory - University of Tennessee, October 2015
- Maimonide seminar - Herzilya, Israel, November 2015.

Marc Shapiro gave the following invited talks:

- Royal Holloway London University, May 2015.
- Universidade Nova de Lisboa, May 2015.
- W. on Chemistry of Concurrent and Distributed Programming II, Agadir, Morocco, May 2015.
- CurryOn, Prague, July 2015.
- Workshop on Large-Scale Distributed Systems (LADIS), Monterey CA USA, Oct. 2015.
- CodeMesh, London, Nov. 2015.

9.1.5. Scientific expertise

Swan Dubois was member of the scientific committee for the assistant professor position n°1642 at University Paris-Sud.

Pierre Sens served on the jury for the "Prix de thèse Gilles Kahn" (SIF - Académie des Sciences).

Pierre Sens also chaired the selection committees for an assistant professor position at Grenoble University.

9.1.6. Research administration

- Pierre is Member of the scientific council of UPMC and officer at UPMC Vice Presidence of Research and Innovation.

9.2. Teaching - Supervision - Juries

9.2.1. 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, Oliver Marin, Sébastien Monnet, Franck Petit, Pierre Sens, Advanced distributed algorithms, M2, UPMC Sorbonne Universités, France

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

Master: Luciana Arantes, Sébastien Monnet, Pierre Sens, Julien Sopena, Operating systems kernel, M1, UPMC Sorbonne Universités, France

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

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

Master: Sébastien Monnet, 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, 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, 22 h, M1, Polytech UPMC Sorbonne Universités, France.

9.2.2. Supervision

PhD: Raluca Diaconu, “Passage à l'échelle pour les mondes virtuels,” UPMC, 01/23/2015, Joaqquin Keller (Orange lab), Sébastien Monnet, Pierre Sens.

PhD: Lokesh Gidra, “Ramasse-miettes pour les machines virtuelles sur les processeurs multicoeurs,” UPMC, 09/28/2015, Gaël Thomas, Marc Shapiro, Julien Sopena.

PhD: Karine Pires, “Diffusion et Transcodage à Grande Échelle de Flux Vidéo en Direct,” UPMC, 03/31/2015, Gwendal Simon, Sébastien Monnet, Pierre Sens.

PhD: Maxime Véron, “Arbitrage décentralisé pour les jeux massivement parrallèles,” UPMC, 09/25/2015, Olivier Marin, Sébastien Monnet, Pierre Sens.

PhD: Marek Zawirski, “Cohérence à terme fiable avec des types de données répliquées,” UPMC, 01/14/2015, Marc Shapiro.

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 : Antoine Blin, "Execution of real-time applications on a small multicore embedded system", since April 2012, Gilles Muller (Whisper) and Julien Sopena, CIFRE Renault

PhD in progress: Marjorie Bournat, “Gathering in robot networks”, UPMC, since Sep. 2014, 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.
PhD in progress: Rudyar Cortes,“Un Environnement à grande échelle pour le traitement de flots massifs de données,” UPMC, funded by Chile government, since Sep. 2013, Olivier Marin, Luciana Arantes, Pierre Sens.
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: Mohamed Hamza Kaaouachi, “Autonomic Distributed Environments for Mobility”, UPMC/Chart-LUTIN (Labex SMART), Franck Petit, Swan Dubois, and François Jouen (Chart).
PhD in progress: Maxime Lorrillere, “A kernel cooperative cache for virtualized environments”, UPMC, Sébastien Monnet, Julien Sopena, Pierre Sens.
PhD in progress: Mahsa Najafzadeh, UPMC, funded by Inria competitive grant (Cordi-S), since Nov. 2012, Marc Shapiro.
PhD in progress: Yoann Péron, “Development of an adaptive recommendation system”, UPMC/Makazi, Franck Petit, Patrick Gallinari, Matthias Oehler (Makazi).
PhD in progress: Alejandro Z. Tomsic, UPMC, funded by SyncFree, since Feb. 2014, Marc Shapiro.
PhD in progress: Guillaume Turchini, “Scalable platform for massively multiplayer online games..”, UPMC, since Sep. 2015, Sébastien Monnet.
PhD in progress: Tao Thanh Vinh, UPMC, CIFRE, since Feb. 2014, Marc Shapiro, Vianney Rancurel (Scality).
PhD in progress: Gauthier Voron, “Big-Os : un OS pour les grands volumes de données,”, UPMC, since Sep. 2014, Gaël Thomas, Pierre Sens.
Master 1 : Mohamed Bekthaoui, “Cliques maximales dans les TVGs”, ENS Lyon, Swan Dubois.

9.2.3. Juries

Franck Petit was the reviewer of:

- T. Langner, PhD ETHZ, Zurich, Suisse. (Advisor: R. Watenhoffer)
- M. Djibril Faye, PhD LIP, ENS Lyon. (Advisor: E. Caron)
- M. Khaled, PhD MIS, Amiens. (Advisors: V. Villain and F. Levé)

Franck Petit was Chair of :

- D. Bonnin, PhD LaBRI, Bordeaux. (Advisors: C. Travers and Y. Métivier)
- M. Véron, PhD LIP6, Paris. (Advisors: P. Sens, O. Marin, and S. Monnet)

Pierre Sens was the reviewer of:

- G. Da Costa, HDR IRIT, Toulouse
- V. Marangozova, HDR LIG, Grenoble
- G. Fedak, HDR LIP, Lyon
- P. Li, PhD Bordeaux (Advisors: R. Namyst, E. Brunnet)
- R. Leroy, PhD LIFL - Maison de la Simulation, (Advisor: N. Melab)
- T. Martsinkevich, PhD LRI (Advisor: F. Cappello)
- M. Antoine, PhD Univ. Nice (Advisor: F. Baude)
- C. Gómez-Calzado, PhD Univ. San Sebastian, Spain (Advisor: M. Larrea)
- J. De la Houssaye, PhD Univ. Évry (Advisor: F. Pommereau)
- F. Zanon Boito, PhD LIG (Advisors: Y. Denneulin, P. Naveaux)
- A. El Rheddane, PhD LIG (Advisor: N. de Palma)

Pierre Sens was Chair of the defense committees of:

- S. Monnet, HDR LIP6, Paris
- D. Conan, HDR Télécom Sud Paris, Evry
- R. Angarita, PhD Dauphine, Paris (Advisor: M. Rukoz)
- X. Han, PhD Télécom Sud Paris, Evry (Advisor: N. Crespi)
- R. Farahbakhsh, Han, PhD Télécom Sud Paris, Evry (Advisor: N. Crespi)
- D. Yang, PhD Télécom Sud Paris, Evry (Advisor: D. Zeghlache)
- H. Xiong, PhD Télécom Sud Paris, Evry (Advisor: D. Zeghlache)
- L. Guo, PhD LIP6, Paris, (Advisor: G. Muller)

Marc Shapiro was a reviewer on the defense committee of Mehdi Ahmed-Nacer (Nancy).

Sébastien Monnet was member of the defense committee of Amadou Diarra, Phd Grenoble university, Grenoble (Advisor: V. Quema).

9.3. Popularization

Sébastien Monnet is responsible for the Science Festival at UPMC for the LIP6.

Sébastien Monnet and Julien Sopena animated an activity during the Science Festival 2015.

10. 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] J. Beauquier, M. Gradinariu Potop-Butucaru, C. Johnen. *Randomized self-stabilizing and space optimal leader election under arbitrary scheduler on rings*, in "Distributed Computing", 2007, vol. 20, n^{o} 1, pp. 75-93

[3] M. Bertier, L. Arantes, P. Sens. *Distributed Mutual Exclusion Algorithms for Grid Applications: A Hierarchical Approach*, in "JPDC: Journal of Parallel and Distributed Computing", 2006, vol. 66, pp. 128–144

[4] B. Ducourthial, S. Khalfallah, F. Petit. *Best-effort group service in dynamic networks*, in "22nd Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA)", 2010, pp. 233-242

[5] 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

[6] 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

[7] R. HU, J. SOPENA, L. ARANTES, P. SENS, I. DEMEURE. *Fair Comparison of Gossip Algorithms over Large-Scale Random Topologies*, in "31th IEEE International Symposium on Reliable Distributed Systems (SRDS'12)", IEEE Computer Society Press, October 2012

[8] S. LEGTCHENKO, S. MONNET, G. THOMAS. *Blue banana: resilience to avatar mobility in distributed MMOGs*, in "The 40th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)", July 2010

[9] J.-P. LOZI, F. DAVID, G. THOMAS, J. L. LAWALL, G. MULLER. *Remote Core Locking: Migrating Critical-Section Execution to Improve the Performance of Multithreaded Applications*, in "USENIX Annual Technical Conference", USENIX, June 2012, pp. 65-76

[10] N. PALIX, G. THOMAS, S. SAHA, C. CALVÈS, J. L. LAWALL, G. MULLER. *Faults in Linux: Ten Years Later*, in "Sixteenth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS 2011)", Newport Beach, CA, USA, March 2011

[11] M. SAEIDA ARDEKANI, T. B. DOUGLAS. *A Self-Configurable Geo-Replicated Cloud Storage System*, in "11th USENIX Symposium on Operating Systems Design and Implementation (OSDI 14)", Broomfield, CO, United States, 2014, https://hal.inria.fr/hal-01102803

[12] 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

[13] 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

[14] 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

[15] N. SCHIPER, P. SUTRA, F. PEDONE. *P-Store: Genuine Partial Replication in Wide Area Networks*, in "Symp. on Reliable Dist. Sys. (SRDS)", New Dehli, India, IEEE Comp. Society, October 2010, pp. 214–224

[16] 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

[17] 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

[18] R. Diaconu. *Scalability for virtual worlds*, Université Pierre et Marie Curie - Paris VI, January 2015, https://tel.archives-ouvertes.fr/tel-01166029

[19] L. Gidra. *Garbage Collector for memory intensive applications on NUMA architectures*, Inria Paris Rocquencourt ; LIP6 - Laboratoire d'Informatique de Paris 6, September 2015, https://tel.archives-ouvertes.fr/tel-01248125

[20] S. Monnet. *Contributions to data replication in large-scale distributed systems*, UPMC Université Paris VI, November 2015, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-01241522

[21] K. Pires. *Delivery and transcoding for large scale live streaming systems*, UPMC Université Paris VI, March 2015, https://tel.archives-ouvertes.fr/tel-01244564

[22] M. P. A. Veron. *Scalable services for massively multiplayer online games*, UPMC Université Paris VI, September 2015, https://hal.archives-ouvertes.fr/tel-01241856

[23] M. Zawirski. *Dependable Eventual Consistency with Replicated Data Types*, Universite Pierre et Marie Curie, January 2015, https://tel.archives-ouvertes.fr/tel-01248051

Articles in International Peer-Reviewed Journals

[24] N. Alon, H. Attiya, S. Dolev, S. Dubois, M. Potop-Butucaru, S. Tixeuil. *Practically stabilizing SWMR atomic memory in message passing systems*, in "Journal of Computer and System Sciences", June 2015, vol. 81, n^{o} 4, pp. 692-701 [*DOI :* 10.1016/j.jcss.2014.11.014], http://hal.upmc.fr/hal-01123697

[25] J. Anjos, I. C. Izurieta, W. Kolberg, A. L. Tibola, L. Arantes, C. Geyer. *MRA++: Scheduling and data placement on MapReduce for heterogeneous environments*, in "Future Generation Computer Systems", January 2015, vol. 42, pp. 22-35 [*DOI :* 10.1016/j.future.2014.09.001], https://hal.archives-ouvertes.fr/hal-01197424

[26] V. Balegas, S. Duarte, C. Ferreira, R. Rodrigues, M. Najafzadeh, M. Shapiro, N. Preguiça. *Towards Fast Invariant Preservation in Geo-replicated Systems*, in "Operating Systems Review", January 2015, vol. 49, n^{o} 1, 5 p. [*DOI :* 10.1145/2723872.2723889], https://hal.inria.fr/hal-01111206

[27] S. Dubois, T. Masuzawa, S. Tixeuil. *Maximum Metric Spanning Tree Made Byzantine Tolerant*, in "Algorithmica", September 2015, vol. 73, n^{o} 1, pp. 166-201 [*DOI :* 10.1007/s00453-014-9913-5], https://hal.archives-ouvertes.fr/hal-01151748

[28] J. Lejeune, L. Arantes, J. Sopena, P. Sens. *A fair starvation-free prioritized mutual exclusion algorithm for distributed systems*, in "Journal of Parallel and Distributed Computing", September 2015, vol. 83, pp. 13-29 [*DOI :* 10.1016/j.jpdc.2015.04.002], https://hal.archives-ouvertes.fr/hal-01178757

[29] H. Senger, V. Gil-Costa, L. Arantes, C. A. Marcondes, M. Marin, L. M. Sato, F. A. B. da Silva. *BSP Cost and Scalability Analysis for MapReduce Operations*, in "Concurrency and Computation: Practice and Experience", October 2015 [*DOI :* 10.1002/cpe.3628], https://hal.inria.fr/hal-01254275

[30] D. Serrano, S. Bouchenak, Y. Kouki, F. Alvares De Oliveira Jr., T. Ledoux, J. Lejeune, J. Sopena, L. Arantes, P. Sens. *SLA guarantees for cloud services*, in "Future Generation Computer

Systems", January 2016, vol. 54, pp. 233–246 [*DOI :* 10.1016/J.FUTURE.2015.03.018], https://hal.archives-ouvertes.fr/hal-01162654

[31] G. SILVESTRE, D. BUFFONI, K. PIRES, S. MONNET, P. SENS. *Boosting Streaming Video Delivery with WiseReplica*, in "Transactions on Large-Scale Data- and Knowledge-Centered Systems", 2015, vol. XX, pp. 34-58 [*DOI :* 10.1007/978-3-662-46703-9_2], https://hal.archives-ouvertes.fr/hal-01198678

Articles in National Peer-Reviewed Journals

[32] M. LORRILLERE, J. SOPENA, S. MONNET, P. SENS. *Conception et évaluation d'un système de cache réparti adapté aux environnements virtualisés*, in "Technique et Science Informatiques", 2015, vol. 34, n^{o} 1-2, 22 p. [*DOI :* 10.3166/TSI.34.101-123], https://hal.inria.fr/hal-01250099

[33] M. VÉRON, O. MARIN, S. MONNET, P. SENS. *Etude des services de matchmaking dans les jeux mutlijoueurs en ligne: récupérer les traces utilisateur afin d'améliorer l'expérience de jeu*, in "Revue des Sciences et Technologies de l'Information - Série TSI : Technique et Science Informatiques", 2015, forthcoming, http://hal.upmc.fr/hal-01158008

International Conferences with Proceedings

[34] K. ALTISEN, A. COURNIER, S. DEVISMES, A. DURAND, F. PETIT. *Élection autostabilisante en un nombre polynomial de pas de calcul*, in "ALGOTEL 2015 — 17èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Beaune, France, June 2015, https://hal.archives-ouvertes.fr/hal-01145472

[35] L. ARANTES, M. BOURNAT, R. FRIEDMAN, O. MARIN, P. SENS. *Elastic Management of Byzantine Faults*, in "11th European Dependable Computing Conference (EDCC 2015)", Paris, France, I. GASHI, Y. BUSNEL (editors), Proceedings of Fast Abstract - EDCC 2015, September 2015, https://hal.archives-ouvertes.fr/hal-01226601

[36] L. ARANTES, R. FRIEDMAN, O. MARIN, P. SENS. *Probabilistic Byzantine Tolerance for Cloud Computing*, in "34th International Symposium on Reliable Distributed Systems (SRDS 2015)", Montreal, Canada, September 2015, http://hal.upmc.fr/hal-01166767

[37] 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

[38] V. BALEGAS, D. SERRA, S. DUARTE, C. FERREIRA, M. SHAPIRO, R. RODRIGUES, N. PREGUIÇA. *Extending Eventually Consistent Cloud Databases for Enforcing Numeric Invariants*, in "Symp. on Reliable Dist. Sys. (SRDS)", Montréal, Canada, Symp. on Reliable Dist. Sys. (SRDS), IEEE Comp. Society, September 2015 [*DOI :* 10.1109/SRDS.2015.32], https://hal.inria.fr/hal-01248192

[39] L. BLIN, F. BOUBEKEUR, S. DUBOIS. *A Self-Stabilizing Memory Efficient Algorithm for the Minimum Diameter Spanning Tree under an Omnipotent Daemon*, in "29rd IEEE International Symposium on Parallel and Distributed Processing (IPDPS 2015)", Hyberabad, India, IEEE, May 2015, pp. 1065-1074 [*DOI :* 10.1109/IPDPS.2015.44], https://hal.archives-ouvertes.fr/hal-01201859

[40] N. BRAUD-SANTONI, S. DUBOIS, M. H. KAAOUACHI, F. PETIT. *A Generic Framework for Impossibility Results in Time-Varying Graphs*, in "17th Workshop on Advances on Parallel and Distributed Processing Sympo-

sium (APDCM'15)", Hyderabad, India, IEEE, May 2015, pp. 483-489 [*DOI :* 10.1109/IPDPSW.2015.59], https://hal.archives-ouvertes.fr/hal-01235800

[41] M. CALLAU-ZORI, L. ARANTES, J. SOPENA, P. SENS. *MERCi-MIsS: Should I turn off my servers?*, in "The 15th IFIP International Conference on Distributed Applications and Interoperable Systems", Grenoble, France, Lecture Notes in Computer Science, Springer International Publishing, June 2015, vol. 9038, pp. 16-29 [*DOI :* 10.1007/978-3-319-19129-4_2], https://hal.archives-ouvertes.fr/hal-01213507

[42] R. CORTÉS, X. BONNAIRE, O. MARIN, P. SENS. *FreeSplit: A Write-Ahead Protocol to Improve Latency in Distributed Prefix Tree Indexing Structures*, in "29th IEEE International Conference on Advanced Information Networking and Applications (AINA-2015)", Gwangju, South Korea, March 2015, http://hal.upmc.fr/hal-01095702

[43] R. CORTÉS, O. MARIN, X. BONNAIRE, L. ARANTES, P. SENS. *A Scalable Architecture for Spatio-Temporal Range Queries over Big Location Data*, in "14th IEEE International Symposium on Network Computing and Applications - IEEE NCA'15", Cambridge, MA, United States, September 2015, http://hal.upmc.fr/hal-01183200

[44] T. CRAIN, M. SHAPIRO. *Designing a causally consistent protocol for geo-distributed partial replication*, in "W. on Principles and Practice of Consistency for Distributed Data (PaPoC)", Bordeaux, France, W. on Principles and Practice of Consistency for Distributed Data (PaPoC), ACM, April 2015 [*DOI :* 10.1145/2745947.2745953], https://hal.inria.fr/hal-01218204

[45] S. DEVISMES, A. LAMANI, F. PETIT, S. TIXEUIL. *Optimal Torus Exploration by Oblivious Robots*, in "NETYS", Agadir, Morocco, Networked Systems - Third International Conference, NETYS 2015, Agadir, Morocco, Springer, May 2015, http://hal.upmc.fr/hal-01131962

[46] 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

[47] S. DUBOIS, M. H. KAAOUACHI, F. PETIT. *Dynamisme et Domination*, in "ALGOTEL 2015 — 17èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Beaune, France, June 2015, https://hal.archives-ouvertes.fr/hal-01145496

[48] S. DUBOIS, M. H. KAAOUACHI, F. PETIT. *Enabling Minimal Dominating Set in Highly Dynamic Distributed Systems*, in "17th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS'15)", Edmonton, Canada, 17th International Symposium, SSS 2015, Edmonton, AB, Canada, August 18-21, 2015, Proceedings, Springer International Publishing, August 2015, vol. 9212, pp. 51-66 [*DOI :* 10.1007/978-3-319-21741-3_4], https://hal.archives-ouvertes.fr/hal-01235826

[49] 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

[50] 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

[51] D. Jeanneau, T. Rieutord, L. Arantes, P. Sens. *A Failure Detector for k -Set Agreement in Dynamic Systems*, in "14th IEEE International Symposium on Network Computing and Applications - (NCA-2015)", Cambridge, United States, 2015, https://hal.inria.fr/hal-01250233

[52] J. Lejeune, L. Arantes, J. Sopena, P. Sens. *Reducing synchronization cost in distributed multi-resource allocation problem*, in "44th International Conference on Parallel Processing", Beijing, China, 44th International Conference on Parallel Processing, September 2015, https://hal.archives-ouvertes.fr/hal-01162329

[53] M. Lorrillere, J. Sopena, S. Monnet, P. Sens. *Puma: pooling unused memory in virtual machines for I/O intensive applications*, in "Proceedings of the 8th ACM International Systems and Storage Conference", Haifa, Israel, ACM, May 2015 [*DOI :* 10.1145/2757667.2757669], https://hal.archives-ouvertes.fr/hal-01154515

[54] T. Rieutord, L. Arantes, P. Sens. *Détecteur de défaillances minimal pour le consensus adapté aux réseaux inconnus*, in "ALGOTEL 2015 — 17èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Beaune, France, June 2015, https://hal.archives-ouvertes.fr/hal-01144111

[55] A. Rosseto, C. Geyer, L. Arantes, P. Sens. *A Failure Detector That Gives Information on the Degree of Confidence in the System*, in "20th IEEE Symposium on Computers and Communication (ISCC-2015)", Larnaca, Cyprus, July 2015, https://hal.archives-ouvertes.fr/hal-01213499

[56] A. Rosseto, C. Rolim, V. Leithardt, G. Borges, C. Geyer, L. Arantes, P. Sens. *A new unreliable failure detector for self-healing in ubiquitous environments*, in "The 29th IEEE International Conference on Advanced Information Networking and Applications (AINA-2015)", Gwangiu, South Korea, March 2015, pp. 316-323 [*DOI :* 10.1109/AINA.2015.201], https://hal.archives-ouvertes.fr/hal-01213333

[57] V. Simon, S. Monnet, M. Feuillet, P. Robert, P. Sens. *Scattering and placing data replicas to enhance long-term durability*, in "The 14th IEEE International Symposium on Network Computing and Applications (IEEE NCA15)", Cambridge, United States, September 2015, 6 p. [*DOI :* 10.1109/NCA.2015.15], https://hal.inria.fr/hal-01206960

[58] V. Tao Thanh, M. Shapiro, V. Rancurel. *Merging Semantics for Conflict Updates in Geo-Distributed File Systems*, in "ACM Int. Systems and Storage Conf. (Systor)", Haifa, Israel, 2015, pp. 10.1–10.12 [*DOI :* 10.1145/2757667.2757683], https://hal.inria.fr/hal-01248190

[59] A. Tomsic, T. Crain, M. Shapiro. *An empirical perspective on causal consistency*, in "W. on Principles and Practice of Consistency for Distributed Data (PaPoC)", Bordeaux, France, ACM (editor), 2015-04-21, ACM, April 2015, vol. 49, n^{o} 1, 15 p. [*DOI :* 10.1145/2745947.2745949], https://hal.inria.fr/hal-01218208

[60] A. Tomsic, P. Sens, J. Garcia, L. Arantes, J. Sopena. *2W-FD: A Failure Detector Algorithm with QoS*, in "The 29th IEEE International Parallel and Distributed Processing Symposium", Hyderabad, India, May 2015, pp. 885-893 [*DOI :* 10.1109/IPDPS.2015.74], https://hal.archives-ouvertes.fr/hal-01213509

[61] G. TURCHINI, S. MONNET, O. MARIN. *Scalability and availability for massively multiplayer online games*, in "11th European Dependable Computing Conference (EDCC 2015)", Paris, France, I. GASHI, Y. BUSNEL (editors), Proceedings of Fast Abstract - EDCC 2015, September 2015, https://hal.archives-ouvertes.fr/hal-01226608

[62] M. VÉRON, O. MARIN, S. MONNET, P. SENS. *RepFD - Using reputation systems to detect failures in large dynamic networks*, in "44th International Conference on Parallel Processing (ICPP-2015)", Beijing, China, September 2015, http://hal.upmc.fr/hal-01150288

[63] 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 (editor), Middleware 2015, December 2015 [*DOI :* 10.1145/2814576.2814733], https://hal.inria.fr/hal-01248194

National Conferences with Proceedings

[64] *Best Paper*
V. GAUTHIER, G. THOMAS, P. SENS, V. QUEMA. *Optimisation mémoire dans une architecture NUMA : comparaison des gains entre natif et virtualisé*, in "Conférence en Parallélisme, Architecture et Système, (COMPAS'15)", Lille, France, 2015, https://hal.inria.fr/hal-01253189.

[65] M. LORRILLERE, J. POUDROUX, J. SOPENA, S. MONNET. *Gestion dynamique du cache entre machines virtuelles*, in "Conférence d'Informatique en Parallélisme, Architecture et Système", Lille, France, Compas'2015, June 2015, pp. 1-10, https://hal.archives-ouvertes.fr/hal-01171226

Conferences without Proceedings

[66] R. HU. *Efficient Probabilistic Information Broadcast Algorithm over Random Geometric Topologies*, in "GLOBECOM", San Diego, United States, December 2015, https://hal.archives-ouvertes.fr/hal-01232676

[67] A. Z. TOMSIC, T. CRAIN, M. SHAPIRO. *Scaling geo-replicated databases to the MEC environment*, in "W. on Planetary-Scale Distributed Systems", Montréal, Canada, 2015, Co-located with SRDS. No proceedings, https://hal.inria.fr/hal-01248195

Research Reports

[68] K. ALTISEN, S. DEVISMES, A. DURAND, F. PETIT. *Gradual Stabilization under τ-Dynamics*, VERIMAG UMR 5104, Université Grenoble Alpes, France ; LIP6 UMR 7606, Inria, UPMC Sorbonne Universités, France, October 2015, https://hal.archives-ouvertes.fr/hal-01215190

[69] S. DUBOIS, M. H. KAAOUACHI, F. PETIT. *Enabling Minimal Dominating Set in Highly Dynamic Distributed Systems*, UPMC Sorbonne Universités/CNRS/Inria - EPI REGAL, January 2015, https://hal.inria.fr/hal-01111610

[70] D. JEANNEAU, T. RIEUTORD, L. ARANTES, P. SENS. *A Failure Detector for k-Set Agreement in Asynchronous Dynamic Systems*, UPMC Sorbonne Universités/CNRS/Inria - EPI REGAL ; Inria, March 2015, n^o RR-8727, https://hal.inria.fr/hal-01151739

[71] J. Lejeune, L. Arantes, J. Sopena, P. Sens. *Reducing synchronization cost in distributed multi-resource allocation problem*, Ecole des Mines de Nantes, Inria, LINA ; Sorbonne Universités, UPMC, CNRS, Inria, LIP6 ; Inria, February 2015, n° RR-8689, https://hal.inria.fr/hal-01120808

[72] D. R. Malikireddy, M. Saeida Ardekani, M. Shapiro. *Emulating Geo-Replication on Grid5000*, Inria – Centre Paris-Rocquencourt ; Inria, April 2015, n° RT-0455, 15 p. , https://hal.inria.fr/hal-01149185

[73] A. G. M. Rossetto, L. Arantes, P. Sens, C. R. Geyer. *Impact: an Unreliable Failure Detector Based on Processes' Relevance and the Confidence Degree in the System*, Université Pierre et Marie Curie ; Inria Paris-Rocquencourt - Regal ; Universidade Federal do Rio Grande do Sul, January 2015, https://hal.inria.fr/hal-01136595

[74] M. Zawirski, N. Preguiça, S. Duarte, A. Bieniusa, V. Balegas, M. Shapiro. *Write Fast, Read in the Past: Causal Consistency for Client-side Applications*, Inria – Centre Paris-Rocquencourt ; Inria, May 2015, n° RR-8729, https://hal.inria.fr/hal-01158370

Scientific Popularization

[75] M. Shapiro. *Bringing the cloud closer to users*, in "EU Research", 2015, vol. 2015, n° 1, pp. 68–69, https://hal.inria.fr/hal-01248193

Other Publications

[76] M. Lorrillere, J. Sopena, S. Monnet, P. Sens. *Puma: pooling unused memory in virtual machines for I/O intensive applications*, May 2015, The 8th ACM International Systems and Storage Conference, Poster, https://hal.archives-ouvertes.fr/hal-01154566

[77] M. Saeida Ardekani, P. Sutra, N. Preguiça, M. Shapiro. *Non-Monotonic Snapshot Isolation*, December 2015, Submitted for publication, https://hal.inria.fr/hal-01248200

[78] M. Shapiro, M. Najafzadeh. *CISE Safety Tool*, October 2015, https://hal.inria.fr/medihal-01242710

[79] M. Zawirski, C. Baquero, A. Bieniusa, N. Preguiça, M. Shapiro. *Eventually Consistent Register Revisited*, November 2015, 7 p. , In order to converge in the presence of concurrent updates, modern eventually consistent replication systems rely on causality information and operation semantics. It is relatively easy to use semantics of high-level operations on replicated data structures, such as sets, lists, etc. However, it is difficult to exploit semantics of operations on registers, which store opaque data. In existing register designs, concurrent writes are resolved either by the application, or by arbitrating them according to their timestamps. The former is complex and may require user intervention, whereas the latter causes arbitrary updates to be lost. In this work, we identify a register construction that generalizes existing ones by combining runtime causality ordering, to identify concurrent writes, with static data semantics, to resolve them. We propose a simple conflict resolution template based on an application-predefined order on the domain of values. It eliminates or reduces the number of conflicts that need to be resolved by the user or by an explicit application logic. We illustrate some variants of our approach with use cases, and how it generalizes existing designs., https://hal.inria.fr/hal-01242700

References in notes

[80] 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