IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Rennes

Université Rennes 1

Activity Report 2013

Project-Team ASAP

As Scalable As Possible: Foundations of Large Scale Dynamic Distributed Systems

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed Systems and middleware

Table of contents

Project-Team ASAP

Keywords: Distributed System, Social Networks, Web Personalization, Theory Of Distributed Computing, Privacy, Big Data, Peer-to-peer

Creation of the Project-Team: 2007 July 01.

1. Members

Research Scientists

Anne-Marie Kermarrec [Team leader, Inria, Senior Researcher, HdR]
Davide Frey [Inria, Researcher]
George Giakkoupis [Inria, Researcher]

Faculty Members

Marin Bertier [INSA Rennes, Associate Professor]
Michel Raynal [Univ. Rennes I, Professor, HdR]
François Taïani [Univ. Rennes I, Professor, HdR]

Engineers

Fabien André [Inria, granted by EIT ICT LABS GA project, from Sept to Nov 2013]
Heverson Borba Ribeiro [Inria, granted by EIT ICT LABS GA project]
Antoine Boutet [Inria, granted by ERC-PoC/ALLYOURS project]
Raziel Carvajal Gomez [Inria, from Oct 2013]
Jacques Falcou [Inria, granted by ERC-PoC/ALLYOURS project, from Apr 2013]
Mathieu Goessens [Inria, granted by EIT ICT LABS GA project, until Jul 2013]
Jean-François Verdonck [Inria, granted by EIT ICT LABS GA project, from Jan 2013]

PhD Students

Mohammad Alaggan [Univ. Rennes I, until Oct 2013]
Resmi Ariyattu Chandrasekharannair [Univ. Rennes I, from Dec 2013]
Tyler Crain [Univ. Rennes I, until Feb 2013]
Fabien André [Technicolor, granted by CIFRE, from Dec 2013]
Stéphane Delbruel [Univ. Rennes I, from Oct 2013]
Ali Gouta [Orange Labs, granted by CIFRE, from Dec 2011]
Arnaud Jégou [Univ. Rennes I, granted by GOSSPLE project, until Oct 2013]
Eleni Kanellou [Univ. Rennes I, until Oct 2013]
Konstantinos Kloudas [Inria, granted by GOSSPLE project, until Mar 2013]
Nupur Mittal [Inria, granted by Google Focused Award Alter-Egos, from Aug 2013]
Javier Olivares [University of Chili, external grant, from Sep 2013]
Antoine Rault [Inria, granted by Conseil Régional de Bretagne, from Oct 2012]
Julien Stainer [Univ. Rennes I, from Oct 2011]
Alexandre Van Kempen [Inria, granted by GOSSPLE project, until Mar 2013]

Post-Doctoral Fellow

Juan Manuel Tirado Martin [Inria, granted by Google Focused Award Alter-Egos, from Dec 2013]

Visiting Scientists

Zahra Aghazadeh [Univ. Calgary , from Jul 2013 until Jul 2013]
Yahya Benkaouz [Inria, from Dec 2013]

Administrative Assistant

Cécile Bouton [Inria, AER]

Others

Yuri Barssi [Univ. Rennes I, Internship granted by EIT ICT LABS, from Sep 2013]

Sylvain Favre [Univ. Rennes I, Internship, from Jul 2013 until Aug 2013]
Hoël Kervadec [Inria, Internship, from Jul 2013 until Sep 2013]
Nabil Kmihi [Inria, Internship, from May 2013 until Oct 2013]
Vincent L Honore [Inria, Internship, from Jun 2013 until Aug 2013]
Olivier Ruas [ENS Cachan, Internship, from Feb 2013 until Jun 2013]
Asiff Shaik [Inria, Internship, until Jan 2013]

2. Overall Objectives

2.1. General objectives

The ASAP Project-Team focuses its research on a number of aspects in the design of large-scale distributed systems. Our work, ranging from theory to implementation, aims to satisfy the requirements of large-scale distributed platforms, namely scalability, and dealing with uncertainty and malicious behaviors. The recent evolutions that the Internet has undergone yield new challenges in the context of distributed systems, namely the explosion of social networking, the prevalence of notification over search, the privacy requirements and the exponential growth of user-generated data introducing more dynamics than ever.

2.1.1. Scalability

The past decade has been dominated by a major shift in *scalability* requirements of distributed systems and applications mainly due to the exponential growth of network technologies (Internet, wireless technology, sensor devices, etc.). Where distributed systems used to be composed of up to a hundred of machines, they now involve thousands to millions of computing entities scattered all over the world and dealing with a huge amount of data. In addition, participating entities are highly dynamic, volatile or mobile. Conventional distributed algorithms designed in the context of local area networks do not scale to such extreme configurations. The ASAP project aims to tackle these *scalability* issues with novel distributed protocols for large-scale dynamic environments.

2.1.2. Personalization

The need for scalability is also reflected in the huge amounts of data generated by Web 2.0 applications. Their fundamental promise, achieving *personalization*, is limited by the enormous computing capacity they require to deliver effective services like storage, search, or recommendation. Only a few companies can afford the cost of the immense cloud platforms required to process users' personal data and even they are forced to use off-line and cluster-based algorithms that operate on quasi-static data. This is not acceptable when building, for example, a large-scale news recommendation platform that must match a multitude of user interests with a continuous stream of news. Scalable algorithms for personalization systems are one of our main research objectives.

2.1.3. Uncertainty

Effective design of distributed systems requires protocols that are able to deal with *uncertainty*. Uncertainty used to be created by the effect of asynchrony and failures in traditional distributed systems, it is now the result of many other factors. These include process mobility, low computing capacity, network dynamics, scale, and more recently the strong dependence on personalization which characterizes user-centric Web 2.0 applications. This creates new challenges such as the need to manage large quantities of personal data in a scalable manner while guaranteeing the privacy of users.

2.1.4. Malicious behaviors and privacy

One particularly important form of uncertainty is associated with faults and *malicious* (or arbitrary) behaviors often modeled as a generic *adversary*. Protecting a distributed system partially under the control of an adversary is a multifaceted problem. On the one hand, protocols must tolerate the presence of participants that may inject spurious information, send multiple information to processes, because of a bug, an external attack, or even an unscrupulous person with administrative access (*Byzantine* behaviors). On the other hand, they must also be able to preserve *privacy* by hiding confidential data from unauthorized participants or from external observers. Within a twenty-year time frame, we can envision that social networks, email boxes, home hard disks, and their online backups will have recorded the personal histories of hundreds of millions of individuals. This raises privacy issues raised by potentially sharing sensitive information with arbitrarily large communities of users.

Successfully managing this scenario requires novel techniques integrating distributed systems, privacy, and data mining with radically different research subjects such as social sciences. In the coming years, we aim to develop these techniques both by building on the expertise acquired during the GOSSPLE project. Gossip algorithms will remain one of the core technologies we use. In these protocols, every node contacts only a few random nodes in each round and exchanges a small amount of information with them. This form of communication is attractive because it offers reasonable performance and is, at the same time, simple, scalable, fault-tolerant, and decentralized. Often, gossip algorithms are designed so that nodes need only little computational power and a small amount of storage space. This makes them perfect candidates to address our objectives: namely dealing with personalization, privacy, and user-generated content, on a variety of devices, including resource-constrained terminals such as mobile phones.

2.2. Structure of the team

Our ambitious goal is to provide the algorithmic foundations of large-scale dynamic distributed systems, ranging from abstractions to real deployment. This is reflected in the following objectives:

2.2.1. Objective 1: Decentralized personalization

Our first objective is to offer full-fledged personalization in notification systems. Today, almost everyone is suffering from an overload of information that hurts both users and content providers. This suggests that not only will notification systems take a prominent role but also that, in order to be useful, they should be personalized to each and every user depending on her activity, operations, posts, interests, etc. In the GOSSPLE implicit instant item recommender, through a simple interface, users get automatically notified of items of interest for them, without explicitly subscribing to feeds or interests. They simply have to let the system know whether they like the items they receive (typically through a like/dislike button). Throughout the system's operation the personal data of users is stored on their own machines, which makes it possible to provide a wide spectrum of privacy guarantees while enabling cross-application benefits.

Our goal here is to provide a fully decentralized solution without ever requiring users to reveal their private preferences.

2.2.2. Objective 2: Personalization: Cloud computing meets p2p

Our second objective is to move forward in the area of **personalization**. Personalization is one of the biggest challenges addressed by most large stake holders.

Hybrid infrastructures for personalisation. So far, social filtering techniques have mainly been implemented on centralized architectures relying on smart heuristics to cope with an increasing load of information. We argue however that, no matter how smart these heuristics and how powerful the underlying machines running them, a fully centralized approach might not be able to cope with the exponential growth of the Internet and, even if it does, the price to be paid might simply not be acceptable for its users (privacy, ecological footprint, etc.).

At the other end of the spectrum, lie fully decentralized systems where the collaborative filtering system is implemented by the machines of the users themselves. Such approaches are appealing for both scalability and privacy reasons. With respect to scalability, storage and computational units naturally grow with the number of users. Furthermore, a p2p system provides an energy-friendly environment where every user can feel responsible for the ecological foot-print of her exploration of the Internet. With respect to privacy, users are responsible for the management of their own profiles. Potential privacy threats therefore do not come from a big-brother but may still arise due to the presence of other users.

We have a strong experience in devising and experimenting with such kinds of p2p systems for various forms of personalization. More specifically, we have shown that personalization can be effective while maintaining a reasonable level of privacy. Nevertheless, frequent connections/disconnections of users make such systems difficult to maintain while addressing privacy attacks. For this reason, we also plan to explore hybrid approaches where the social filtering is performed by the users themselves, as in a p2p manner, whereas the management of connections-disconnections, including authentication, is managed through a server-based architecture. In particular, we plan to explore the trade-off between the quality of the personalization process, its efficiency and the privacy guarantees.

2.2.3. Objective 3: Privacy-aware decentralized computations

Gossip algorithms have also been studied for more complex global tasks, such as computation of network statistics or, more generally, aggregation functions of input values of the nodes (e.g., sum, average, or max). We plan to pursue this research direction both from a theoretical and from a practical perspective. We provide two examples of these directions below.

Computational capabilities of gossip. On the theoretical side, we have recently started to study gossip protocols for the assignment of unique IDs from a small range to all nodes (known as the *renaming* problem) and computing the rank of the input value of each node. We plan to further investigate the class of global tasks that can be solved efficiently by gossip protocols.

Private computations on decentralized data. On a more practical track, we aim to explore the use of gossip protocols for decentralized computations on privacy sensitive data. Recent research on private data bases, and on homomorphic encryption, has demonstrated the possibility to perform complex operations on encrypted data. Yet, existing systems have concentrated on relatively small-scale applications. In the coming years, we instead plan to investigate the possibility to build a framework for querying and performing operations for large-scale decentralized data stores. To achieve this, we plan to disseminate queries in an epidemic fashion through a network of data sources distributed on a large scale while combining privacy preserving techniques with decentralized computations. This would, for example, enable the computation of statistical measures on large quantities of data without needing to access and disclose each single data item.

2.2.4. Objective 4: Information dissemination over social networks

While we have been studying information dissemination in practical settings (such as WhatsUp in GOSSPLE), modelling such dynamic systems is still in its infancy. We plan to complement our practical work on gossip algorithms and information dissemination along the following axes:

Rumour spreading is a family of simple randomized algorithms for information dissemination, in which nodes contact (uniformly) random neighbours to exchange information with them. Despite their simplicity these protocols have proved very efficient for various network topologies. We are interested in studying their properties in specific topologies such as social networks be they implicit (interest-based as in GOSSPLE) or explicit (where users choose their friends as in Facebook). Recently, there has been some work on bounding the speed of rumour spreading in terms of abstract properties of the network graph, especially the graph's expansion properties of conductance and vertex expansion. It has been shown that high values for either of these guarantees fast rumour spreading—this should be related to empirical observations that social networks have high expansion. Some works established increasingly tighter upper bounds for rumour spreading in term of conductance or vertex expansion, but these bounds are not tight.

Our objective is to prove the missing tight upper bound for rumour spreading with vertex expansion. It is known that neither conductance nor vertex expansion are enough by themselves to completely characterize the speed of rumour spreading: are there graphs with bad expansion in which rumours spread fast? We plan to explore more refined notions of expansion and possibly other abstract graph properties, to establish more accurate bounds. Another interesting and challenging problem is the derivation of general lower bounds for rumour spreading as a function of abstract properties of graphs. No such bounds are currently known.

Overcoming the dependence on expansion: Rumour spreading algorithms have very nice properties as their simplicity, good performances for many networks but they may have very poor performance for some networks, even though these networks have small diameter, and thus it is possible to achieve fast information dissemination with more sophisticated protocols. Typically nodes may choose the neighbours to contact with some non-uniform probabilities that are determined based on information accumulated by each node during the run of the algorithm. These algorithms achieve information dissemination in time that is close to the diameter of the network. These algorithms, however, do not meet some of the other nice properties of rumour spreading, most importantly, robustness against failures. We are investigating algorithms that combine the good runtime of these latest protocols with the robustness of rumour spreading. Indeed these algorithms assumed that the network topology does not change during their run. In view of the dynamism of real networks, in which nodes join and leave and connection between nodes change constantly, we have to address dynamic network models. We plan to investigate how the classic rumour spread algorithms perform in the face of changes. We plan also in this area to reduce the size of the messages they use, which can be high even if the amount of useful information that must be disseminated is small.

Competing rumours: Suppose now that two, or more, conflicting rumours (or opinions) spread in the network, and whenever a node receives different rumours it keeps only one of them. Which rumour prevails, and how long does it take until this happens? Similar questions have been studied in other contexts but not in the context of rumour spreading. The *voter* model is a well studied graph process that can be viewed as a competing rumour process that follows the classic PULL rumour spreading algorithm. However, research has only recently started to address the question of how long it takes until a rumour prevails. An interesting variant of the problem that has not been considered before is when different rumours are associated with different weights (some rumours are more convincing than others). We plan to study the above models and variations of them, and investigate their connection to the standard rumour spreading algorithms. This is clearly related to the dissemination of news and personalization in social networks.

2.2.5. Objective 5: Computability and efficiency of distributed systems

A very relevant challenge (maybe a Holy Grail) lies in the definition of a computation model appropriate to dynamic systems. This is a fundamental question. As an example there are a lot of peer-to-peer protocols but none of them is formally defined with respect to an underlying computing model. Similarly to the work of Lamport on "static" systems, a model has to be defined for dynamic systems. This theoretical research is a necessary condition if one wants to understand the behavior of these systems. As the aim of a theory is to codify knowledge in order it can be transmitted, the definition of a realistic model for dynamic systems is inescapable whatever the aim we have in mind, be it teaching, research or engineering.

Distributed computability: Among the fundamental theoretical results of distributed computing, there is a list of problems (e.g., consensus or non-blocking atomic commit) that have been proved to have no deterministic solution in asynchronous distributed computing systems prone to failures. In order such a problem to become solvable in an asynchronous distributed system, that system has to be enriched with an appropriate oracle (also called failure detector). We have been deeply involved in this research and designed optimal consensus algorithms suited to different kind of oracles. This line of research paves the way to rank the distributed computing problems according to the "power" of the additional oracle they required (think of "additional oracle" as "additional assumptions"). The ultimate goal would be the statement of a distributed computing hierarchy, according to the minimal assumptions needed to solve distributed computing problems (similarly to the Chomsky's hierarchy that ranks problems/languages according to the type of automaton they need to be solved).

Distributed computing abstractions: Major advances in sequential computing came from machine-independent data abstractions such as sets, records, etc., control abstractions such as while, if, etc., and modular constructs such as functions and procedures. Today, we can no longer envisage not to use these abstractions. In the "static" distributed computing field, some abstractions have been promoted and proved to be useful. Reliable broadcast, consensus, interactive consistency are some examples of such abstractions. These abstractions have well-defined specifications. There are both a lot of theoretical results on them (mainly decidability and lower bounds), and numerous implementations. There is no such equivalent for dynamic distributed systems, i.e. for systems characterized by nodes that may join and leave, or that may change their characteristics at runtime. Our goal is to define such novel abstractions, thereby extending the theory of distributed systems to the dynamic case.

2.3. Highlights of the Year

- Google Focused Award (only 2 in Europe this year).
- ERC Proof of Concept Grant AllYours (2013)
- New associate team with the University of Calgary, Canada (RADCON).
- Michel Raynal published two new books on Concurrent Programming and Distributed Algorithms.
- Anne-Marie Kermarrec was elected the Vice Chair of *ACM Eurosys (ACM European Chapter of SIGOPS).*
- Anne-Marie Kermarrec is a member of *Academia Europea* since Sept 2013.
- Michel Raynal has been nominated adjunct professor of University of Hong Kong.

3. Research Program

3.1. Distributed computing

Distributed computing [1] was born in the late seventies when people started taking into account the intrinsic characteristics of physically distributed systems. The field then emerged as a specialized research area distinct from networks, operating systems and parallelism. Its birth certificate is usually considered as the publication in 1978 of Lamport's most celebrated paper "*Time, clocks and the ordering of events in a distributed system*" [60] (that paper was awarded the Dijkstra Prize in 2000). Since then, several high-level journals and (mainly ACM and IEEE) conferences have been devoted to distributed computing. The distributed systems area has continuously been evolving, following the progresses of all the above-mentioned areas such as networks, computing architecture, operating systems.

The last decade has witnessed significant changes in the area of distributed computing. This has been acknowledged by the creation of several conferences such as NSDI and IEEE P2P. The NSDI conference is an attempt to reassemble the networking and system communities while the IEEE P2P conference was created to be a forum specialized in peer-to-peer systems. At the same time, the EuroSys conference originated as an initiative of the European Chapter of the ACM SIGOPS to gather the system community in Europe.

3.2. Theory of distributed systems

Finding models for distributed computations prone to asynchrony and failures has received a lot of attention. A lot of research in this domain focuses on what can be computed in such models, and, when a problem can be solved, what are its best solutions in terms of relevant cost criteria. An important part of that research is focused on distributed computability: what can be computed when failure detectors are combined with conditions on process input values for example. Another part is devoted to model equivalence. What can be computed with a given class of failure detectors? Which synchronization primitives is a given failure class equivalent to? These

[1] This is an extract from Michel Raynal's new book [42].

are among the main topics addressed in the leading distributed computing community. A second fundamental issue related to distributed models, is the definition of appropriate models suited to dynamic systems. Up to now, the researchers in that area consider that nodes can enter and leave the system, but do not provide a simple characterization, based on properties of computation instead of description of possible behaviors [61], [55], [56]. This shows that finding dynamic distributed computing models is today a "Holy Grail", whose discovery would allow a better understanding of the essential nature of dynamic systems.

3.3. Peer-to-peer overlay networks

A standard distributed system today is related to thousands or even millions of computing entities scattered all over the world and dealing with a huge amount of data. This major shift in scalability requirements has lead to the emergence of novel computing paradigms. In particular, the peer-to-peer communication paradigm imposed itself as the prevalent model to cope with the requirements of large scale distributed systems. Peer-to-peer systems rely on a symmetric communication model where peers are potentially both clients and servers. They are fully decentralized, thus avoiding the bottleneck imposed by the presence of servers in traditional systems. They are highly resilient to peers arrivals and departures. Finally, individual peer behavior is based on a local knowledge of the system and yet the system converges toward global properties.

A peer-to-peer overlay network logically connects peers on top of IP. Two main classes of such overlays dominate, structured and unstructured. The differences relate to the choice of the neighbors in the overlay, and the presence of an underlying naming structure. Overlay networks represent the main approach to build large-scale distributed systems that we retained. An overlay network forms a logical structure connecting participating entities on top of the physical network, be it IP or a wireless network. Such an overlay might form a structured overlay network [62], [63], [64] following a specific topology or an unstructured network [59], [65] where participating entities are connected in a random or pseudo-random fashion. In between, lie weakly structured peer-to-peer overlays where nodes are linked depending on a proximity measure providing more flexibility than structured overlays and better performance than fully unstructured ones. Proximity-aware overlays connect participating entities so that they are connected to close neighbors according to a given proximity metric reflecting some degree of affinity (computation, interest, etc.) between peers. We extensively use this approach to provide algorithmic foundations of large-scale dynamic systems.

3.4. Epidemic protocols

Epidemic algorithms, also called gossip-based algorithms [58], [57], constitute a fundamental topic in our research. In the context of distributed systems, epidemic protocols are mainly used to create overlay networks and to ensure a reliable information dissemination in a large-scale distributed system. The principle underlying technique, in analogy with the spread of a rumor among humans via gossiping, is that participating entities continuously exchange information about the system in order to spread it gradually and reliably. Epidemic algorithms have proved efficient to build and maintain large-scale distributed systems in the context of many applications such as broadcasting [57], monitoring, resource management, search, and more generally in building unstructured peer-to-peer networks.

3.5. Malicious process behaviors

When assuming that processes fail by simply crashing, bounds on resiliency (maximum number of processes that may crash), number of exchanged messages, number of communication steps, etc. either in synchronous and augmented asynchronous systems (recall that in purely asynchronous systems some problems are impossible to solve) are known. If processes can exhibit malicious behaviors, these bounds are seldom the same. Sometimes, it is even necessary to change the specification of the problem. For example, the consensus problem for correct processes does not make sense if some processes can exhibit a Byzantine behavior and thus propose an arbitrary value. In this case, the validity property of consensus, which is normally "a decided value is a proposed value", must be changed to "if all correct processes propose the same value then only this value can be decided." Moreover, the resilience bound of less than half of faulty processes is at least lowered to "less than a third of Byzantine processes." These are some of the aspects that underlie our studies in the context of the classical model of distributed systems, in peer-to-peer systems and in sensor networks.

3.6. Online social networks

Social Networks have rapidly become a fundamental component of today's distributed applications. Web 2.0 applications have dramatically changed the way users interact with the Internet and with each other. The number of users of websites like Flickr, Delicious, Facebook, or MySpace is constantly growing, leading to significant technical challenges. On the one hand, these websites are called to handle enormous amounts of data. On the other hand, news continue to report the emergence of privacy threats to the personal data of social-network users. Our research aims to exploit our expertise in distributed systems to lead to a new generation of scalable, privacy-preserving, social applications.

4. Application Domains

4.1. Overview

The results of the research targeted in ASAP span a wide range of applications. Below are a few examples.

- Personalized web search.
- Recommendation.
- Social networks.
- Notification systems.
- Distributed storage.
- Video streaming.

5. Software and Platforms

5.1. MediEgo: A recommendation solution for webmasters

Participants: Antoine Boutet, Jacques Falcou, Arnaud Jégou, Anne-Marie Kermarrec, Jean-François Verdonck.

Contact:	Anne-Marie Kermarrec
Licence:	Proprietary
Presentation:	Recommendation solution for webmasters
Status:	Beta version, IDDN.FR.001.490030.000.S.P.2013.000.30000 on 09/12/2013

MEDIEGO is a solution for content recommendation based on the users navigation history. The solution 1) collects the usages of the Web users and store them in a profile; 2) uses this profile to associate to each user her most similar users; 3) leverages this implicit network of close users in order to infer their preferences and recommend advertisements and recommendations. MEDIEGO achieves scalability using a sampling method, which provides very good results at a drastically reduced cost.

5.2. MediEgo Dashboard: A personalized news dashboard

Participants: Yuri Barssi, Antoine Boutet, Anne-Marie Kermarrec, Jean-François Verdonck.

Contact:	Antoine Boutet
Licence:	Proprietary
Status:	Beta version

This work has led to the development of MEDIEGO Dashboard, a personalized news recommendation system. In MEDIEGO Dashboard, users benefit from a personalized stream of news matching their interests. Additionally, users can use explicit subscriptions as well as post content and navigate through tags. MEDIEGO Dashboard is available through a web interface and a mobile-based Android application. To provide personalization, MEDIEGO Dashboard exploits the users' opinions regarding their received news to identify users with similar interests. MEDIEGO Dashboard is centralized and it allows us to test and evaluate different recommendation schemes. In collaboration with EIT/ICT Lab, an experiment has been conducted with a set of users at Trento (Italie). This experiment allowed us to collect traces and to perform a user survey to assess and improve our solution. This solution will soon be interconnected to AllYours-P2P.

5.3. AllYours-P2P: A distributed news recommender (former WhatsUp)

Participants: Heverson Borba Ribeiro, Antoine Boutet, Davide Frey, Arnaud Jégou, Anne-Marie Kermarrec, Jean-François Verdonck.

Contact:	Antoine Boutet
Licence:	AGPL 3.0
Presentation:	A distributed news recommender
Status:	Beta version, IDDN.FR.001.500002.000.S.P.2013.000.30000 on 09/12/2013

In the context of the AllYours EIT/ICT Labs project, we refined the implementation of WhatsUp into the AllYours-P2P application. The application provides a distributed recommendation system aimed to distribute instant news in a large scale dynamic system. It consists of two parts, running on each peer: an embedded application server, based on Jetty, and a web interface accessible from any web browser. The application server exchanges information with other peers in the system, while the web interface displays news items and collects the opinions of the user.

5.4. HyRec: A hybrid recommender system

Participants: Antoine Boutet, Davide Frey, Anne-Marie Kermarrec.

Contact:	Antoine Boutet
Licence:	Proprietary
Status:	Beta version, IDDN.FR.001.500007.000.S.P.2013.000.30000 on 09/12/2013

This work leads to the development of HyRec, a hybrid recommender system. The motivation of this work is to explore solutions that could in some sense democratize personalization by making it accessible to any content provider company without generating huge investments. HyRec implements a user-based collaborative filtering scheme and offloads CPU-intensive recommendation tasks to front-end client browsers, while retaining storage and orchestration tasks within back-end servers. HyRec seeks to provide the scalability of P2P approaches without forcing content providers to give up the control of the system.

5.5. GossipLib: A library for gossip-based applications

Participants: Heverson Borba Ribeiro, Davide Frey, Anne-Marie Kermarrec.

Contact:	Heverson Borba Ribeiro, Davide Frey
Licence:	AGPL 3.0
Presentation:	Library for gossip protocols
Status:	Alpha version, IDDN.FR.001.500001.000.S.P.2013.000.10000 on 09/12/2013

GossipLib is a library consisting of a set of JAVA classes aimed to facilitate the development of gossip-based application in a large-scale setting. It provides developers with a set of support classes that constitute a solid starting point for building any gossip-based application. GossipLib is designed to facilitate code reuse and testing of distributed application, and provides also the implementation of a number of standard gossip protocols that may be used out of the box or extended to build more complex protocols and applications. These include for example the peer-sampling protocols for overlay management. GossipLib also provides facility for the configuration and deployment of applications as final-product but also as research prototype in environments like PlanetLab, clusters, network emulators, and even as event-based simulation. The code developed with GossipLib can be run both as a real application and in simulation.

5.6. YALPS: A library for P2P applications

Participants: Heverson Borba Ribeiro, Davide Frey, Anne-Marie Kermarrec.

Contact:	Heverson Borba Ribeiro, Davide Frey
Licence:	Open Source
Presentation:	Library for p2p applications
Status:	Beta version, IDDN.FR.001.500003.000.S.P.2013.000.10000 on 09/12/2013

YALPS is an open-source Java library designed to facilitate the development, deployment, and testing of distributed applications. Applications written using YALPS can be run both in simulation and in real-world mode without changing a line of code or even recompiling the sources. A simple change in a configuration file will load the application in the proper environment. A number of features make YALPS useful both for the design and evaluation of research prototypes and for the development of applications to be released to the public. Specifically, YALPS makes it possible to run the same application as a simulation or in a real deployment. Applications communicate by means of application-defined messages which are then routed either through UDP/TCP or through YALPS's simulation infrastructure. In both cases, YALPS's communication layer offers features for testing and evaluating distributed protocols and applications. Communication channels can be tuned to incorporate message losses or to constrain their outgoing bandwidth. Finally, YALPS includes facilities to support operation in the presence of NATs and firewalls using relaying and NAT-traversal techniques. This work was done in collaboration with Maxime Monod (EPFL).

5.7. HEAP: Heterogeneity-aware gossip protocol

Participants: Davide Frey, Arnaud Jégou, Anne-Marie Kermarrec.

Contact:	Davide Frey
Licence:	Open Source
Presentation:	Java Application
Status:	Release & ongoing development

This work has been done in collaboration with Vivien Quéma (CNRS Grenoble), Maxime Monod and Rachid Guerraoui (EPFL), and has lead to the development of a video streaming platform based on HEAP, *HEterogeneity-Aware gossip Protocol*. The platform is particularly suited for environment characterized by heterogeneous bandwidth capabilities such as those comprising ADSL edge nodes. HEAP is, in fact, able to dynamically leverage the most capable nodes and increase their contribution to the protocol, while decreasing by the same proportion that of less capable nodes. During the last few months, we have integrated HEAP with the ability to dynamically measure the available bandwidth of nodes, thereby making it independent of the input of the user.

6. New Results

6.1. Models and abstractions for distributed systems

6.1.1. Randomized loose renaming in O(loglog n) time

Participant: George Giakkoupis.

Renaming is a classic distributed coordination task in which a set of processes must pick distinct identifiers from a small namespace. In [24], we consider the time complexity of this problem when the namespace is linear in the number of participants, a variant known as loose renaming. We give a non-adaptive algorithm with $O(\log\log n)$ (individual) step complexity, where n is a known upper bound on contention, and an adaptive algorithm with step complexity $O((\log\log k)^2)$, where k is the actual contention in the execution. We also present a variant of the adaptive algorithm which requires $O(k\log\log k)$ *total* process steps. All upper bounds hold with high probability against a strong adaptive adversary. We complement the algorithms with an $\Omega(\log\log n)$ expected time lower bound on the complexity of randomized renaming using test-and-set operations and linear space. The result is based on a new coupling technique, and is the first to apply to non-adaptive randomized renaming. Since our algorithms use $O(n)$ test-and-set objects, our results provide matching bounds on the cost of loose renaming in this setting.

This work was done in collaboration with Dan Alistarh, James Aspnes, and Philipp Woelfel.

6.1.2. An O(sqrt n) space bound for obstruction-free leader election

Participant: George Giakkoupis.

In [32] we present a deterministic obstruction-free implementation of leader election from $O(\sqrt{n})$ atomic $O(\log n)$-bit registers in the standard asynchronous shared memory system with n processes. We provide also a technique to transform any deterministic obstruction-free algorithm, in which any process can finish if it runs for b steps without interference, into a randomized wait-free algorithm for the oblivious adversary, in which the expected step complexity is polynomial in n and b. This transformation allows us to combine our obstruction-free algorithm with the leader election algorithm by Giakkoupis and Woelfel (2012), to obtain a fast randomized leader election (and thus test-and-set) implementation from $O(\sqrt{n})$ $O(\log n)$-bit registers, that has expected step complexity $O(\log^* n)$ against the oblivious adversary. Our algorithm provides the first sub-linear space upper bound for obstruction-free leader election. A lower bound of $\Omega(\log n)$ has been known since 1989 (Styer and Peterson, 1989). Our research is also motivated by the long-standing open problem whether there is an obstruction-free consensus algorithm which uses fewer than n registers.

This work was done in collaboration with Maryam Helmi, Lisa Higham, and Philipp Woelfel.

6.1.3. Broadcast in recurrent dynamic systems

Participants: Michel Raynal, Julien Stainer.

This work [50] proposes a simple broadcast algorithm suited to dynamic systems where links can repeatedly appear and disappear. The algorithm is proved correct and a simple improvement is introduced, that reduces the number and the size of control messages. As it extends in a simple way a classical network traversal algorithm (due to A. Segall, 1983) to the dynamic context, the proposed algorithm has also pedagogical flavor.

This work has been done in collaboration with Jiannong Cao and Weigang Wu.

6.1.4. Computing in the presence of concurrent solo executions

Participants: Michel Raynal, Julien Stainer.

In a wait-free model any number of processes may crash. A process runs solo when it computes its local output without receiving any information from other processes, either because they crashed or they are too slow. While in wait-free shared-memory models at most one process may run solo in an execution, any number of processes may have to run solo in an asynchronous wait-free message-passing model. This work [47] is on the computability power of models in which several processes may concurrently run solo. We introduced a family of round-based wait-free models, called the d-solo models, $1 \le d \le n$, where up to d processes may run solo. Then we gave a characterization of the colorless tasks that can be solved in each d-solo model. We also introduced the (d,ϵ)-solo approximate agreement task, which generalizes ϵ-approximate agreement, and proves that (d,ϵ)-solo approximate agreement can be solved in the d-solo model, but cannot be solved in the $(d+1)$-solo model. We also studied the relation linking d-set agreement and (d,ϵ)-solo approximate agreement in asynchronous wait-free message-passing systems. These results establish for the first time a hierarchy of wait-free models that, while weaker than the basic read/write model, are nevertheless strong enough to solve non-trivial tasks.

This work was done in collaboration with Maurice Herlihy and Sergio Rajsbaum.

6.1.5. Relating message-adversaries and failure detectors

Participants: Michel Raynal, Julien Stainer.

A message adversary is a daemon that suppresses messages in round-based message-passing synchronous systems in which no process crashes. A property imposed on a message adversary defines a subset of messages that cannot be eliminated by the adversary. It has recently been shown that when a message adversary is constrained by a property denoted TOUR (for tournament), the corresponding synchronous system and the asynchronous crash-prone read/write system have the same computability power for task solvability. In this work [39] we introduced new message adversary properties (denoted SOURCE and QUORUM), and shown that the synchronous round-based systems whose adversaries are constrained by these properties are characterizations of classical asynchronous crash-prone systems (1) in which processes communicate through atomic read/write registers or point-to-point message-passing, and (2) enriched with failure detectors such as Ω and Σ. Hence these properties characterize maximal adversaries, in the sense that they define strongest message adversaries equating classical asynchronous crash-prone systems. They consequently provide strong relations linking round-based synchrony weakened by message adversaries with asynchrony restricted with failure detectors. This not only enriches our understanding of the synchrony/asynchrony duality, but also allows for the establishment of a meaningful hierarchy of property-constrained message adversaries.

6.1.6. A hierarchy of agreement problems from simultaneous consensus to set agreement

Participants: Michel Raynal, Julien Stainer.

In this work [38] we investigated the relation linking the s-simultaneous consensus problem and the k-set agreement problem in wait-free message-passing systems. To this end, we defined the (s,k)-SSA problem which captures jointly both problems: each process proposes a value, executes s simultaneous instances of a k-set agreement algorithm, and has to decide a value so that no more than sk different values are decided. We also introduced a new failure detector class denoted $Z_{s,k}$, which is made up of two components, one focused on the "shared memory object" that allows the processes to cooperate, and the other focused on the liveness of (s,k)-SSA algorithms. A novelty of this failure detector lies in the fact that the definition of its two components are intimately related. We designed a $Z_{s,k}$-based algorithm that solves the (s,k)-SSA problem, and shown that the "shared memory"-oriented part of $Z_{s,k}$ is necessary to solve the (s,k)-SSA problem (this generalizes and refines a previous result that showed that the generalized quorum failure detector Σ_k is necessary to solve k-set agreement). We finally, investigated the structure of the family of (s,k)-SSA problems and introduced generalized (asymmetric) simultaneous set agreement problems in which the parameter k can differ in each underlying k-set agreement instance. Among other points, it shows that, for $s,k > 1$, (a) the $(sk,1)$-SSA problem is strictly stronger that the (s,k)-SSA problem which is itself strictly stronger than the $(1,ks)$-SSA problem, and (b) there are pairs (s_1,k_1) and (s_2,k_2) such that $s_1k_1 = s_2k_2$ and (s_1,k_1)-SSA and (s_2,k_2)-SSA are incomparable.

6.2. Large-scale and user-centric distributed systems

6.2.1. FreeRec: An anonymous and distributed personalization architecture

Participants: Antoine Boutet, Davide Frey, Arnaud Jégou, Anne-Marie Kermarrec, Heverson Borba Ribeiro.

FreeRec is an anonymous decentralized peer-to-peer architecture designed to bring personalization while protecting the privacy of its users [17], [30], [44]. FreeRec's decentralized approach makes it independent of any entity wishing to collect personal data about users. At the same time, its onion-routing-like gossip-based overlay protocols effectively hide the association between users and their interest profiles without affecting the quality of personalization. The core of FreeRec consists of three layers of overlay protocols: the bottom layer, rps, consists of a standard random peer sampling protocol ensuring connectivity; the middle layer, PRPS, introduces anonymity by hiding users behind anonymous proxy chains, providing mutual anonymity; finally, the top clustering layer identifies for each anonymous user, a set of anonymous nearest neighbors. We demonstrate the effectiveness of FreeRec by building a decentralized and anonymous content dissemination system. Our evaluation by simulation, our PlanetLab experiments, and our probabilistic analysis show that FreeRec effectively decouples users from their profiles without hampering the quality of personalized content delivery.

6.2.2. HyRec: A hybrid recommender system

Participants: Antoine Boutet, Davide Frey, Anne-Marie Kermarrec.

The ever-growing amount of data available on the Internet calls for personalization. Yet, the most effective personalization schemes, such as those based on collaborative filtering (CF), are notoriously resource greedy. HyRec is an online cost-effective scalable system for CF personalization. HyRec relies on a hybrid architecture, offloading CPU-intensive recommendation tasks to front-end client browsers, while retaining storage and orchestration tasks within back-end servers. HyRec has been fully implemented and extensively evaluated on several workloads from MovieLens and Digg. We convey the ability of HyRec to significantly reduce the operation costs of the content provider by up to 70% and drastically improve the scalability by up to 500%, with respect to a centralized (or cloud-based recommender approach), while preserving the quality of the personalization. We also show that HyRec is virtually transparent to the users and induces only 3% of the bandwidth consumption of a P2P solution.

6.2.3. Social market

Participants: Davide Frey, Arnaud Jégou, Anne-Marie Kermarrec, Michel Raynal, Julien Stainer.

The ability to identify people that share one's own interests is one of the most interesting promises of the Web 2.0 driving user-centric applications such as recommendation systems or collaborative marketplaces. To be truly useful, however, information about other users also needs to be associated with some notion of trust. Consider a user wishing to sell a concert ticket. Not only must she find someone who is interested in the concert, but she must also make sure she can trust this person to pay for it. Social Market (SM) solves this problem by allowing users to identify and build connections to other users that can provide interesting goods or information and that are also reachable through a trusted path on an explicit social network like Facebook. This year, we extended the contributions presented in 2011, by introducing two novel distributed protocols that combine interest-based connections between users with explicit links obtained from social networks ala Facebook. Both protocols build trusted multi-hop paths between users in an explicit social network supporting the creation of semantic overlays backed up by social trust. The first protocol, TAPS2, extends our previous work on *TAPS* (Trust-Aware Peer Sampling), by improving the ability to locate trusted nodes. Yet, it remains vulnerable to attackers wishing to learn about trust values between arbitrary pairs of users. The second protocol, PTAPS (*Private TAPS*), improves TAPS2 with provable privacy guarantees by preventing users from revealing their friendship links to users that are more than two hops away in the social network. In addition to proving this privacy property, we evaluate the performance of our protocols through event-based simulations, showing significant improvements over the state of the art. In addition to our previous publication on this topic, our recent work led to a paper that appeared in TCS [20].

6.2.4. Privacy-preserving P2P collaborative filtering

Participants: Davide Frey, Anne-Marie Kermarrec, Antoine Rault, François Taïani.

The huge amount of information available at any time in our connected society calls for a mechanism to filter it efficiently. Recommendation systems provide such a mechanism by personalizing the information displayed for each user. However, the collection of personal information by recommendation systems threatens the privacy of users. We address the two needs for recommendation and privacy through a peer-to-peer user-based collaborative filtering system. Recommendation is done ala GOSSPLE by building an overlay network which connects users with similar interests via clustering and random peer sampling. This overlay network is then used to make recommendations based on what similar users liked. Users' privacy is protected in two ways. Users are protected from a Big Brother adversary by the peer-to-peer design of the system in which profiles are stored only by their owners. Users are protected from other malicious users who would try to learn the content of their profiles by our landmark-based cosine similarity measure. It indirectly computes the similarity of two users by comparing their respective similarities with a set of randomly generated profiles, called landmarks. Thus, users can compute their similarity without revealing their profile, contrarily to the regular cosine similarity when used in a peer-to-peer system.

6.2.5. Gossip protocols for renaming and sorting

Participants: George Giakkoupis, Anne-Marie Kermarrec.

In [33] we devise efficient gossip-based protocols for some fundamental distributed tasks. The protocols assume an n-node network supporting point-to-point communication, and in every round, each node exchanges information of size $O(\log n)$ bits with (at most) one other node. We first consider the *renaming* problem, that is, to assign distinct IDs from a small ID space to all nodes of the network. We propose a renaming protocol that divides the ID space among nodes using a natural push or pull approach, achieving logarithmic round complexity with ID space $\{1, \cdots, (1+\epsilon)n\}$, for any fixed $\epsilon > 0$. A variant of this protocol solves the *tight* renaming problem, where each node obtains a unique ID in $\{1, \cdots, n\}$, in $O(\log^2 n)$ rounds. Next we study the following *sorting* problem. Nodes have consecutive IDs 1 up to n, and they receive numerical values as inputs. They then have to exchange those inputs so that in the end the input of rank k is located at the node with ID k. Jelasity and Kermarrec (2006) suggested a simple and natural protocol, where nodes exchange values with peers chosen uniformly at random, but it is not hard to see that this protocol requires $\Omega(n)$ rounds. We prove that the same protocol works in $O(\log^2 n)$ rounds if peers are chosen according to a non-uniform power law distribution.

This work has been done in collaboration with Philipp Woelfel.

6.2.6. Adaptive streaming

Participants: Ali Gouta, Anne-Marie Kermarrec.

HTTP Adaptive Streaming (HAS) is gradually being adopted by Over The Top (OTT) content providers. In HAS, a wide range of video bitrates of the same video content are made available over the internet so that clients' players pick the video bitrate that best fit their bandwidth. Yet, this affects the performance of some major components of the video delivery chain, namely CDNs or transparent caches since several versions of the same content compete to be cached. We investigated the benefits of a Cache Friendly HAS system (CF-DASH), which aims to improve the caching efficiency in mobile networks and to sustain the quality of experience of mobile clients. We presented a set of observations we made on large number of clients requesting HAS contents [34], [35]. Then, we evaluated CF-dash based on trace-driven simulations and testbed experiments. Our validation results are promising. Simulations on real HAS traffic show that we achieve a significant gain in hit-ratio that ranges from 15% up to 50%.

Work was done in collaboration with Yannick Le Louedec, Zied Aouini and Diallo Mamadou.

6.2.7. DynaSoRe: Efficient in-memory store for social applications

Participant: Arnaud Jégou.

Social network applications are inherently interactive, creating a requirement for processing user requests fast. To enable fast responses to user requests, social network applications typically rely on large banks of cache servers to hold and serve most of their content from the cache. The objective of this work is to build a memory cache system for social network applications that optimizes data locality while placing user views across the system. We call this system DynaSoRe (Dynamic Social stoRe). DynaSoRe storage servers monitor access traffic and bring data frequently accessed together closer in the system to reduce the processing load across cache servers and network devices. Our simulation results considering realistic data center topologies show that DynaSoRe is able to adapt to traffic changes, increase data locality, and balance the load across the system. The traffic handled by the top tier of the network connecting servers drops by 94% compared to a static assignment of views to cache servers while requiring only 30% additional memory capacity compared to the whole volume of cached data.

This work was conducted in collaboration with Xiao Bai, Flavio Junqueira, and Vincent Leroy. The product of this collaboration led to the publication of a paper at the Middleware 2013 conference [26].

6.2.8. Adaptive metrics on distributed recommendation systems

Participants: Anne-Marie Kermarrec, François Taïani, Juan Manuel Tirado Martin.

Current distributed recommendation systems are metric based. This means that recommendation quality depends on a single user comparison function. This is a simple solution that cannot cover the particularities of each system. Classically computing intensive data-mining methods have been used in the field of recommendation. However, they are not proper in distributed scenarios due to the lack of a global vision and the existing restrictions in terms of computing power. In this project, we study how to provide and model ad-hoc similarity metrics that can be automatically adapted to a different number of scenarios. We study our solution from two different points of view: recommendation and performance. In the first, we evaluate the capacity of data mining technics to give users relevant recommendations. Second, by exploring the performance of different approaches in order to obtain relevant recommendations we plan to study the trade-off between relevant recommendations and computational cost.

6.2.9. Cliff-Edge Consensus: Agreeing on the precipice

Participants: Michel Raynal, François Taïani.

In this project, we worked on a new form of consensus that allows nodes to agree locally on the extent of crashed regions in networks of arbitrary size. One key property of our algorithm is that it shows local complexity, i.e. its cost is independent of the size of the complete system, and only depends on the shape and extent of the crashed region to be agreed upon. In [40], we motivate the need for such an algorithm, formally define this new consensus problem, propose a fault-tolerant solution, and prove its correctness.

This work was done in collaboration with Geoff Coulson and Barry Porter.

6.2.10. Clustered network coding

Participants: Fabien André, Anne-Marie Kermarrec, Konstantinos Kloudas, Alexandre Van Kempen.

Modern storage systems now typically combine plain replication and erasure codes to reliably store large amount of data in datacenters. Plain replication allows a fast access to popular data, while erasure codes, e.g. Reed-Solomon codes, provide a storage-efficient alternative for archiving less popular data. Although erasure codes are now increasingly employed in real systems, they experience high overhead during maintenance, i.e. upon failures, typically requiring files to be decoded before being encoded again to repair the encoded blocks stored at the faulty node.

In this work, we propose a novel erasure code system, tailored for networked archival systems. The efficiency of our approach relies on a combination of the use of random codes coupled with a clever yet simple clustered placement strategy. Our repair protocol leverages network coding techniques to reduce by 50% the amount of data transferred during maintenance, as several cluster files are repaired simultaneously. We demonstrate both through an analysis and extensive experimental study conducted on a public testbed that our approach dramatically decreases both the bandwidth overhead during the maintenance process and the time to repair data lost upon failure.

This has been done in collaboration with Erwan le Merrer, Nicolas, Le Scouarnec and Gilles Straub.

7. Bilateral Contracts and Grants with Industry

7.1. Technicolor

Participants: Fabien André, Anne-Marie Kermarrec.

We have a contract with Technicolor for collaboration on large-scale infrastructure for recommendation systems . In this context, Anne-Marie Kermarrec is the PhD advisor of Fabien André since November 2013. Fabien André will work on efficient algorithms for heterogeneous data on large-scale platforms.

7.2. Orange Labs

Participants: Ali Gouta, Anne-Marie Kermarrec.

We have had a contract with Orange Labs for collaboration on peer-assisted approaches for caching and recommendation in streaming applications. In this context, Anne-Marie Kermarrec has been the PhD advisor of Ali Gouta since 2012.

7.3. Web Alter-Egos Google Focused Award

Participants: George Giakkoupis, Anne-Marie Kermarrec, Nupur Mittal, Javier Olivares.

Duration: Sep. 2013 - Sep. 2015; Coordinator: Inria and EPFL.

This project addresses the problem of extracting the alter-egos of a Web user, namely profiles of like-minded users who share similar interests, across various Internet applications, in real time and in the presence of high dynamics. Beyond their intrinsic social interest, the profiles of alter-egos of a user are crucial to identify a personalized slice of the Internet that can be leveraged to personalize the Web navigation of that user. The expected outcome of the project is a generic architecture of a Web-Alter-Ego service that can run on various devices and use, as well as be used for, various Web applications.

8. Partnerships and Cooperations

8.1. National initiatives

8.1.1. LABEX CominLabs

Participants: Anne-Marie Kermarrec, Davide Frey, Michel Raynal, François Taïani.

ASAP participates in the CominLabs initiative sponsored by the "Laboratoires d'Excellence" program. The initiative federates the best teams from Bretagne and Nantes regions in the broad area of telecommunications, from electronic devices to wide area distributed applications "over the top." These include, among the others, the Inria teams: ACES, ALF, ASAP, CELTIQUE, CIDRE, DISTRIBCOM, MYRIADS, TEMICS, TEXMEX, and Visages. The scope of CominLabs covers research, education, and innovation. While being hosted by academic institutions, CominLabs builds on a strong industrial ecosystem made of large companies and competitive SMEs. In this context, ASAP received funding for DeSeNt (a collaborative project with the Université de Nantes / LINA).

8.1.2. ANR ARPÈGE project Streams

Participants: Marin Bertier, Michel Raynal.

The Streams project started in November 2010. Beside the ASAP group, it includes teams from Inria Nancy and PARIS. Its aim it to design a real-time collaborative platform based on a peer-to-peer network. For this it is necessary to design a support architecture that offers guarantees on the propagation, security and consistency of the operations and the updates proposed by the different collaborating sites.

8.1.3. ANR project SocioPlug

Participants: Anne-Marie Kermarrec, Davide Frey, Michel Raynal, François Taïani.

SocioPlug is a collaborative ANR project involving Inria (ASAP team), the université de Nantes, and LIRIS (INSA Lyon and Universite Claude Bernard Lyon). The project emerges from the observation that the features offered by the Web 2.0 or by social media do not come for free. Rather they bring the implicit cost of privacy. Users are more of less consciously selling personal data for services. SocioPlug aims to provide an alternative for this model by proposing a novel architecture for large-scale, user centric applications. Instead of concentrating information of cloud platforms owned by a few economic players, we envision services made possible by cheap low-end plug computers available in every home or workplace. This will make it possible to provide a high amount of transparency to users, who will be able to decide their own optimal balance between data sharing and privacy.

8.1.4. DeSceNt CominLabs

Participants: Resmi Ariyattu Chandrasekharannair, Davide Frey, Michel Raynal, François Taïani.

The DeSceNt project aims to ease the writing of distributed programs on a federation of plug computers. Plug computers are a new generation of low-cost computers, such as Raspberry pi (25$), VIA- APC (49$), and ZERO Devices Z802 (75$), which offer a cheap and readily available infrastructure to deploy domestic on-line software. Plug computers open the opportunity for everyone to create cheap nano-clusters of domestic servers, host data and services and federate these resources with their friends, colleagues, and families based on social links. More particularly we will seek in this project to develop novel decentralized protocols than can encapsulate the notion of privacy-preserving federation in plug-based infrastructures. The vision is to use these protocols to provide a programming toolkit that can support the convergent data types being developed by our partner GDD (Grande Données Distribuées) at Université de Nantes.

8.1.5. ANR Blanc project Displexity

Participants: George Giakkoupis, Anne-Marie Kermarrec, Michel Raynal.

The Displexity project started in October 2011. The aim of this ANR project that also involves researchers from Paris and Bordeaux is to establish the scientific foundations for building up a consistent theory of computability and complexity for distributed computing. One difficulty to be faced by DISPLEXITY is to reconcile two non necessarily disjoint sub-communities, one focusing on the impact of temporal issues, while the other focusing on the impact of spatial issues on distributed algorithms.

8.2. European initiatives

8.2.1. FP7 projects

8.2.1.1. ALLYOURS ERC Proof of Concept

Title: AllYours: A distributed privacy-aware instant item recommender

Type: IDEAS

Instrument: ERC Proof of Concept Grant (Starting)

Duration: January 2013 - December 2013.

Coordinator: Inria (France)

See also: http://www.gossple.fr

Abstract: The goal of this PoC proposal is to boost the creation of a start-up (AllYours/MEDIEGO) targeting both Internet users as well as small to medium companies (SME) offering full-fledged personalization in notification systems. AllYours is a direct outcome from the GOSSPLE ERC Starting Grant, and more specifically from one of the activities conducted within the project, that today involves most of the team and forces. In the GOSSPLE ERC SG project, we have invented the concept of implicit social network, built and maintained in a fully decentralized manner so that each user is in charge of her own personalized data, addressing both the privacy concern that users may have with respect to Big Brother-like companies, and scalability as the resources present at the edges of the Internet can then be fully leveraged. The GOSSPLE social network has been the basis of several Web 2.0 applications in order to personalize Web functionalities within the project, such as search, recommendation, query expansion, top-k queries, etc. More specifically, we have been applying the GOSSPLE social network to personalized notification, defining on top of it a novel dissemination protocol. This is P2P-AllYours currently under development. Our MEDIEGO software is now applied to centralized systems for recommendations.

8.2.1.2. TOWARD THE ALLYOURS START-UP

Title: TOWARD THE ALLYOURS START-UP: Focus on the mobile version

Type: EIT-ICT Labs

Instrument: ACLD Computing in the Cloud

Duration: January 2013 - December 2013.

Coordinator: Inria (France)

Partners: Trento Rise, BDP EIT-ICT

See also: http://www.gossple.fr

Abstract: The goal of the Activity proposal is to turn the inventions from the ERC Starting Grant Project GOSSPLE to innovation by setting up a start-up (AllYours) targeting both Internet users as well as small to medium companies (SME) offering full-fledged personalization in notification systems. This proposal focuses on the mobile versions of AllYours software. While the wired setting is a goal of the foreseen startup, this proposal will focus on the mobile versions of E-AllYours and P2P AllYours that will be experimented on the live platform provided by the TrentoRise partners.

8.2.1.3. ERC SG Gossple

Title: GOSSPLE

Type: IDEAS

Instrument: ERC Starting Grant

Duration: September 2008 - August 2013

Coordinator: Inria (France)

See also: http://www.gossple.fr

Abstract: Anne-Marie Kermarrec is the principal investigator of the GOSSPLE ERC starting Grant (Sept. 2008 - Sept. 2013). GOSSPLE aims at providing a radically new approach to navigating the digital information universe. This project has been granted a 1.250.000 euros budget for 5 years.

GOSSPLE aims at radically changing the navigation on the Internet by placing users affinities and preferences at the heart of the search process. Complementing traditional search engines, GOSSPLE will turn search requests into live data to seek the information where it ultimately is: at the user. GOSSPLE precisely aims at providing a fully decentralized system, self-organizing, able to discover, capture and leverage the affinities between users and data.

8.2.2. Collaborations in European programs, except FP7

8.2.2.1. Transform Marie Curie Initial Training Network (ITN)

Participants: Tyler Crain, Eleni Kanellou, Anne-Marie Kermarrec, Michel Raynal.

Program: Marie Curie Initial Training Network

Project acronym: Transform

Project title: Theoretical Foundations of Transactional Memory

Duration: May 2010 - October 2013

Grant agreement no.: 238639

Date of approval of Annex I by Commission: May 26, 2009

Coordinators: Michel Raynal - Panagiota Fatourou

Other partners: Foundation for Research and Technology Hellas ICS FORTH Greece, University of Rennes I UR1 France, Ecole Polytechnique Federale de Lausanne EPFL Switzerland, Technische Universitaet Berlin TUB Germany, and Israel Institute of Technology Technion.

Abstract: Transform is a Marie Curie Initial Training Networks European project devoted to the Theoretical Foundations of Transactional Memory (Major chip manufacturers have shifted their focus from trying to speed up individual processors into putting several processors on the same chip. They are now talking about potentially doubling efficiency on a 2x core, quadrupling on a 4x core and so forth. Yet multi-core is useless without concurrent programming. The constructors are now calling for a new software revolution: the concurrency revolution. This might look at first glance surprising for concurrency is almost as old as computing and tons of concurrent programming models and languages were invented. In fact, what the revolution is about is way more than concurrency alone: it is about concurrency for the masses. The current parallel programming approach of employing locks is widely considered to be too difficult for any but a few experts. Therefore, a new paradigm of concurrent programming is needed to take advantage of the new regime of multicore computers. Transactional Memory (TM) is a new programming paradigm which is considered by most researchers as the future of parallel programming. Not surprisingly, a lot of work is being devoted to the implementation of TM systems, in hardware or solely in software. What might be surprising is the little effort devoted so far to devising a sound theoretical framework to reason about the TM abstraction. To understand properly TM systems, as well as be able to assess them and improve them, a rigorous theoretical study of the approach, its challenges and its benefits is badly needed. This is the challenging research goal undertaken by this MC-ITN. Our goal through this project is to gather leading researchers in the field of concurrent computing over Europe, and combine our efforts in order to define what might become the modern theory of concurrent computing. We aim at training a set of Early Stage Researchers (ESRs) in this direction and hope that, in turn, these ESRs will help Europe become a leader in concurrent computing. Its keywords are Transactional Memory, Parallelization Mechanisms, Parallel Programming Abstractions, Theory, Algorithms, Technological Sciences

8.2.3. Collaborations with major European organizations

Ecole Polytechnique Federale de Lausanne EPFL Switzerland; collaboration on the ERC SG GOSSPLE and Transform, and the Google Focused Award Web-Alter-Egos.

Foundation for Research and Technology Hellas ICS FORTH Greece; collaboration on Transform

8.3. International initiatives

8.3.1. Inria associate teams

8.3.1.1. RADCON

Title: Randomized Algorithms for Distributed Computing and Networks

Inria principal investigator: George Giakkoupis

International Partner:

University of Calgary (Canada) - Department of Computer Science - Philipp Woelfel

Duration: 2013 - 2015

See also: http://www.irisa.fr/asap/radcon

Over recent years, computing systems have seen a massive increase in parallelism and interconnectivity. Peer-to-peer systems, ad-hoc networks, sensor networks, or the "cloud" are based on highly connected and volatile networks. Individual nodes such as cell phones, desktop computers or high performance computing systems rely on parallel processing power achieved through multiple processing units. To exploit the power of massive networks or multiple processors, algorithms must cope with the scale and asynchrony of these systems, and their inherent instability, e.g., due to node, link, or processor failures. In this research project we explore randomized algorithms for large-scale networks of distributed systems, and for shared memory multi-processor systems. For large-scale networks, decentralized gossip protocols have emerged as a standard approach to achieving fault-tolerant communication between nodes with simple and scalable algorithms. We will devise new gossip protocols for various complex distributed tasks, and we will explore the power and limits of gossip protocols in various settings. For shared memory systems, randomized algorithms have proved extremely useful to deal with asynchrony and failures. Sometimes probabilistic algorithms provide the only solution to a problem; sometimes they are more efficient; sometimes they are simply easier to implement. We will devise efficient algorithms for some of the fundamental problems of shared memory computing, such as mutual exclusion, renaming, and consensus.

8.3.2. Inria international partners

University of Calgary

Universidad Nacional Autonoma de Mexico

University of Glasgow

8.3.3. Participation in international programs

8.3.3.1. Demdyn: Inria/CNPq Collaboration

Participants: Marin Bertier, Michel Raynal.

The aim of this project is to exploit dependable aspects of dynamic distributed systems such as VANETs, WiMax, Airborn Networks, DoD Global Information Grid, P2P, etc. Applications that run on these kind of networks have a common point: they are extremely dynamic both in terms of the nodes that take part of them and available resources at a given time. Such dynamics results in instability and uncertainty of the environment which provide great challenges for the implementation of dependable mechanisms that ensure the correct work of the system. This requires applications to be adaptive, for instance, to less network bandwidth or degraded Quality-of-Service (QoS). Ideally, in these highly dynamic scenarios, adaptiveness characteristics of applications should be self-managing or autonomic. Therefore, being able to detect the occurrence of partitions and automatically adapting the applications for such scenarios is an important dependable requirement for such new dynamic environments.

8.4. International research visitors

8.4.1. Visits of international scientists

Zarah Aghazadeh, University of Calgary, from 6 to 27 July 2013

Laurent Fournier, Cup Foundation Toulouse, 13 November 2013

Roy Friedman, Technion University Tel Aviv, 20 December 2013

Christian Grothoff, TU Munich, 26 November 2013

Jean-Loup Guillaume, Lip6 Paris, 12 March 2013

Gilles Tredan, Laass Toulouse, from 2 to 10 September 2013

Philipp Woelfel, University of Calgary, from 6 to 14 July 2013

8.4.2. Internships

Hoël Kervadec; 1 July 2013 to 6 September 2013. "Construction décentralisée de topologies informatiques réparties à mémoire de forme." Supervised by François Taïani.

Nabil Kmihi; 16 May 2013 to 8 November 2013. "Offline social networks." Supervised by Anne-Marie Kermarrec.

Olivier Ruas; 1 February 2013 to 30 June 2013. "Degree-based routing in small world networks." Supervised by George Giakkoupis and Anne-Marie Kermarrec.

Sylvain Fabre; 1 July 2013 to 31 August 2013. "Elaboration d'un overlay qui prend en compte la localisation des noeuds." Supervised by Marin Bertier.

Vincent L'Honore; 10 July 2013 to 2 August 2013. "The development of the mobile-based Android application of MEDIEGO Dashboard." Supervised by Antoine Boutet.

Yahya Benkaouz; 1 December 2013 to 28 February 2014. "La conception et la mise en œuvre d'un réseau social décentralisé respectant la vie privée des utilisateurs et son intégration dans le système Gossple." Supervised by Anne-Marie Kermarrec.

8.4.3. Visits to international teams

George Giakkoupis visited University of Calgary, Canada, twice, from 15 April to 6 May and from 23 November to 14 December. He also visited MPI, Saarbrücken, Germany, from 3 to 9 August.

Anne-Marie Kermarrec was a part-time visiting professor at EPFL, Lausanne.

Anne-Marie Kermarrec visited University of Sydney and NICTA, Australia, Jan 2014.

9. Dissemination

9.1. Scientific animation

- Antoine Boutet was a member of the shadow program committee of *ACM Eurosys'13*, Prague, Chech Republic, Apr. 2013.
- Davide Frey was a program co-chair of the *International Conference on Distributed Computing and Networking (ICDCN), Distributed Computing Track*, Mumbai, India, Jan. 2013.
- Davide Frey was a program committee member of the conferences:
  - *IEEE International Parallel & Distributed Processing Symposium (IPDPS)*, Boston, Massachusetts USA, May 2013.
  - *IEEE International Conference on Peer-to-Peer Computing (P2P)*, Trento, Italy, Sep. 2013.
  - *ACM/IFIP/USENIX International Conference on Middleware (Middleware)*, Beijing, China, Dec. 2013.
  - *IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS)*, Florence, Italy, June 2013.
- George Giakkoupis was a program committee member of the *International Workshop on Foundations of Mobile Computing (FOMC)*, Jerusalem, Israel, Oct. 2013.
- Anne-Marie Kermarrec is a member of the *Academia Europea* since 2013.
- Anne-Marie Kermarrec is a member of the scientific committee of the *Société Informatique de France*.
- Anne-Marie Kermarrec is a member of the *Inria scientific board (Bureau du comité des projets)* in Rennes.
- Anne-Marie Kermarrec was a member of *ASF 2013*.
- Anne-Marie Kermarrec is the chair of the *ACM Software System Award Committee* in 2014.
- Anne-Marie Kermarrec is a member of the *ACM Software System Award Committee* since 2009.
- Anne-Marie Kermarrec is a member of the *ERC panel for Consolidator Grants* since 2013.
- Anne-Marie Kermarrec was the chair of the hiring committee for the position of *Assistant Professor at ENS Lyon (Chaire Inria/ENS)*.
- Anne-Marie Kermarrec was a member of the 2013 hiring committees of: Inria CR, Max Planck Institute for Software Systems, University of Lugano.
- Anne-Marie Kermarrec is a member of the *IEEE Internet Computing Editorial Board*.
- Anne-Marie Kermarrec is the vice-chair of *ACM Eurosys* since Apr. 2013.
- Anne-Marie Kermarrec is/was a program committee member of the conferences:
  - *ACM/IFIP/USENIX International Conference on Middleware (Middleware)*, China, Dec. 2013.
  - *International Conference on Very Large Data Bases (VLDB)*, Trento, Italy, Aug. 2013.
  - *International Conference on Networked Systems (NETYS)*, Marrakech, May 2013.
  - *ACM International Conference on Data Management (SIGMOD)*, NYC, USA, June 2013.
  - *International Conference on Distributed Computing Systems (ICDCS)*, Philadelphia, USA, July 2013.

  - *International Conference on Peer-to-peer systems (P2P)*, Trento, Italy, Sep. 2013.
  - *International Conference on Distributed Computing Systems (ICDCS)*, Madrid, July 2014.
  - *International Conference on Extending Database Technology (EDBT)*, Athens, Greece, Mar. 2014.
  - *ACM Eurosys*, Amsterdam, Apr. 2014.
  - *ACM International Systems and Storage Conference (SYSTOR)*, Haifa, Israel, June 2014.
  - *International Conference on Networked Systems (NETYS)*, Marrakech, May 2014.
- Anne-Marie Kermarrec is a member of the steering committees of *ACM Eurosys*, *Middleware*, and the *Winter School on Hot Topics in Distributed Systems*.
- Anne-Marie Kermarrec is a member of the technical committee of the *ACM Conference on Online Social Networks*.
- Anne-Marie Kermarrec co-organized the *Workshop Inria/Technicolor on Distributed Storage Systems*, Rennes, Nov. 2013.
- Michel Raynal was in the editorial board of the journals: *IEEE Transactions on Computers*, *Journal of Parallel and Distributed Computing*, *Journal of Computer Systems Science and Engineering*, *Foundations of Computing and Decision Sciences*.
- Michel Raynal was a program co-chair of the *International Conference on Distributed Computing and Networking (ICDCN), Distributed Computing Track*, Mumbai, India, Jan. 2013.
- Michel Raynal was a steering committee member of the conferences:
  - *IEEE International Conference on Distributed Computing Systems (ICDCS)*, Philadelphia, USA, July 2013.
  - *International Conference on Distributed Computing and Networking (ICDCN)*, Mumbai, India, Jan. 2013.
- Michel Raynal was a program committee member of the conferences:
  - *ACM Symposium on Principles of Distributed Computing (PODC)*, Montreal, QC, Canada, July 2013.
  - *International Symposium on Distributed Computing (DISC)*, Jerusalem, Israel, Oct. 2013.
- Michel Raynal was an *Adjunct Professor* of the Polytechnic University (PolyU), Hong-Kong.
- Michel Raynal was a *European Representative in the IEEE Technical Committee on Distributed Computing*.
- Michel Raynal was an international liaison co-chair of *ICDCS 2013*.
- François Taïani was as program committee co-chair of the *IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS)*, Florence, Italy, June 2013.
- François Taïani was a program committee member of the conferences:
  - *International Conference on Distributed Computing Systems (ICDCS)*, Philadelphia, USA, July 2013.
  - *Latin-American Symposium on Dependable Computing (LADC)*, Rio de Janeiro, Brazil, Apr. 2013
  - *Dependable and Adaptive Distributed Systems (DADS) Track of the ACM Symposium on Applied Computing*, Coimbra, Portugal, Mar. 2013
- Juan Manuel Tirado was a Program Committee Member of the *Workshop on Middleware for Next Generation Internet Computing (MW4NG)*, Beijing, China, Dec. 2013.

9.2. Invited talks

- George Giakkoupis. *Rumor spreading on dynamic graphs*. CS Theory Seminar, University of Calgary, Canada, Nov. 29 2013.
- George Giakkoupis. *Gossip protocols for information dissemination and other tasks*. Workshop on Storage and Cloud Computing (WOS3), Rennes, France, Nov. 22 2013.
- George Giakkoupis. *Tight bounds for rumor spreading with graph expansion*. PODC Social Network Workshop, Montreal, Canada, Jul. 21 2013.
- George Giakkoupis. *Randomized loose renaming in* $O(\log \log n)$ *time*. ASAP Workshop on Distributed Computing, Rennes, France, Jun. 28 2013.
- George Giakkoupis. *Rumor spreading and graph expansion*. 3rd Pacific Northwest Theory Days Workshop, University of Victoria, Canada, May 5 2013.
- George Giakkoupis. *Rumor spreading and vertex expansion*. Workshop on Epidemic Algorithms and Processes: From Theory to Applications, Dagstuhl, Germany, Jan. 21 2013.
- Mathieu Goessens. *P2P technologies for the Web: Challenges and perspectives*. Mozilla Summit 2013, Brussels, Belgium, Oct. 5 2013.
- Anne-Marie Kermarrec. Tutorial: *Systems support for news recommendation: From P2P to hybrid architectures*. Transform Summer School on Research Directions in Distributed Computing. Heraklion, Crete. June 10-14, 2013.
- Anne-Marie Kermarrec. *Privacy-aware micro-blogging*. University of Sydney and NICTA, Australia, Jan. 2014.
- Michel Raynal. Tutorial: *Checkpointing and debugging in distributed computing*. Latin-American Conference on Dependable Computing (LADC), Rio de Janeiro, Brazil, Apr. 2013.
- Michel Raynal. *Concurrency-related distributed recursion.* International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS), Osaka, Japan, Nov. 2013.
- Michel Raynal. *Hybrid concurrent objects.* DISC workshop on Software Transactional memory, Jerusalem, Israel, Oct. 2013.
- François Taïani. *Deconstructing complex distributed platforms: A report from the trenches*. International Workshop on Automated Specification and Verification of Web Systems (WWV), Florence, Italy, June 2013.
- François Taïani. Workshop La politique des données personnelles: Big data ou contrôle individuel? Institut Rhône-Alpin des Systèmes Complexes (IXXI) and ENS Lyon, Lyon, Nov. 2013.

9.3. Teaching, supervision, and juries

9.3.1. Teaching

- Licence (bachelor) courses:
  - François Taïani, Introduction to Operating Systems, 16h, L3, ESIR / Université of Rennes I, France.
  - Davide Frey, Invited lectures at ENS Cachan, Paris, and ENS Cachan Antenne de Bretagne.
- Master courses:
  - Davide Frey, Scalable Distributed Systems, 10h, M2, Université de Rennes I, France.
  - François Taïani, Synchronization and Parallel Programming, 62h, M1, ESIR / Université de Rennes I, France.
  - François Taïani, Distributed Systems, 48h, M2, ESIR / Université de Rennes I, France.
  - François Taïani, Programming Technologies for the Cloud, 32h, M2, Université de Rennes I, France.
- Doctoral courses:

  - Anne-Marie Kermarrec, Large-scale Systems and Applications, EPFL, Switzerland.

9.3.2. Supervision

- PhD: Mohammad Alaggan. *Private Peer-to-peer similarity computation in personalized collaborative platforms.* Dec 2013. Supervised by Anne-Marie Kermarrec (and Sébastien Gambs).
- PhD: Antoine Boutet. *Décentralisation des systèmes de personnalisation.* Mar 2013. Supervised by Anne-Marie Kermarrec.
- PhD: Tyler Crain. *Faciliter l'utilisation des mémoires transactionnelles logicielles.* Mar 2013. Supervised by Michel Raynal.
- PhD: Konstantinos Kloudas. *Leveraging Content Properties to Optimize Distributed Storage Systems*. Mar 2013. Supervised by Anne-Marie Kermarrec.
- PhD: Alexandre Van Kempen. *Optimiser l'utilisation de la bande passante dans les systèmes de stockage distribué.* Mar 2013. Supervised by Anne-Marie Kermarrec.
- PhD in progress: Fabien André. *Calcul incrémental sur des plates-formes à large échelle.* Started Nov 2013. Supervised by Anne-Marie Kermarrec.
- PhD in progress: Resmi Ariyattu Chandrasekharannair. *Towards Decentralized Federations for Plug-based Decentralized Social Networks*. Started Dec 2013. Supervised by François Taïani.
- PhD in progress: Stéphane Delbruel. *Towards a Decentralized Embryomorphic Storage System*. Started Oct 2013. Supervised by François Taïani.
- PhD in progress: Ali Gouta. *Caching et distribution adaptatifs dans les réseaux de contenus de grande échelle*. Started Dec 2011. Supervised by Anne-Marie Kermarrec.
- PhD in progress: Arnaud Jégou. *Explicit and implicit social networks, harnessing their power through decentralisation*. Started Oct 2010. Supervised by Anne-Marie Kermarrec and Davide Frey. S
- PhD in progress: Nupur Mitall. *Infrastructure et algorithmes pour la recommandation de contenus cross application*. Started Dec 2013. Supervised by George Giakkoupis and Anne-Marie Kermarrec.
- PhD in progress: Javier Olivares. *Web-Alter Ego Platform : A Peer-sourcing infrastructure for personalized privacy-aware event processing*. Started Oct 2013. Supervised by Anne-Marie Kermarrec.
- PhD in progress: Julien Stainer. *Calculabilité distribuée*. Started Oct 2011. Supervised by Michel Raynal.
- PhD in progress: Antoine Rault. *Privacy Through Decentralization*. Started Oct 2012. Supervised by Anne-Marie Kermarrec and Davide Frey.

9.3.3. Juries

- Davide Frey was a member of the PhD jury for Konstantinos Kloudas. *Leveraging Content Properties to Optimize Distributed Storage Systems*. Université de Rennes I, March 6, 2013. (examinateur)
- François Taïani was a member of the following PhD juries:
  - Konstantinos Kloudas *Leveraging Content Properties to Optimize Distributed Storage Systems*. Université de Rennes I / Inria Rennes, March 6, 2013. (examinateur)
  - François Fouquet. *Kevoree : Model@Runtime pour le développement continu de systèmes adaptatifs distribués hétérogènes*. Université de Rennes I / Inria Rennes, March 6, 2013. (examinateur)
  - Jimmy Lauret. *Prévention et détection des interférences inter-aspects : méthode et application à l'aspectisation de la tolérance aux fautes*. Institut National Polytechnique de Toulouse (INP Toulouse) / LAAS-CNRS, May 15, 2013. (rapporteur)
  - Yousri Kouki. *Approche dirigée par les contrats de niveaux de service pour la gestion de l'élasticité du "nuage"*. École nationale supérieure des mines de Nantes, LINA, Dec. 12, 2013. (rapporteur)

10. Bibliography

Major publications by the team in recent years

[1] M. BERTIER, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, V. LEROY. *The Gossple Anonymous Social Network*, in "ACM/IFIP/USENIX 11th International Middleware Conference", India Bangalore, November 2010, http://hal.inria.fr/inria-00515693/en

[2] J. CAO, M. RAYNAL, X. YANG, W. WU. *Design and Performance Evaluation of Efficient Consensus Protocols for Mobile Ad Hoc Networks*, in "IEEE Transactions on Computers", 2007, vol. 56, n^{o} 8, pp. 1055–1070

[3] A. CARNEIRO VIANA, S. MAAG, F. ZAIDI. *One step forward: Linking Wireless Self-Organising Networks Validation Techniques with Formal Testing approaches*, in "ACM Computing Surveys", 2009, http://hal.inria.fr/inria-00429444/en/

[4] R. FRIEDMAN, A. MOSTEFAOUI, S. RAJSBAUM, M. RAYNAL. *Distributed agreement problems and their connection with error-correcting codes*, in "IEEE Transactions on Computers", 2007, vol. 56, n^{o} 7, pp. 865–875

[5] A. J. GANESH, A.-M. KERMARREC, E. LE MERRER, L. MASSOULIÉ. *Peer counting and sampling in overlay networks based on random walks*, in "Distributed Computing", 2007, vol. 20, n^{o} 4, pp. 267-278

[6] G. GIAKKOUPIS, P. WOELFEL. *A tight RMR lower bound for randomized mutual exclusion*, in "STOC - 44th ACM Symposium on Theory of Computing", New York, United States, May 2012, http://hal.inria.fr/hal-00722940

[7] G. GIAKKOUPIS, P. WOELFEL. *On the time and space complexity of randomized test-and-set*, in "PODC - 31st Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing", Madeira, Portugal, July 2012, http://hal.inria.fr/hal-00722947

[8] M. JELASITY, S. VOULGARIS, R. GUERRAOUI, A.-M. KERMARREC, M. VAN STEEN. *Gossip-Based Peer Sampling*, in "ACM Transactions on Computer Systems", August 2007, vol. 41, n^{o} 5

[9] B. MANIYMARAN, M. BERTIER, A.-M. KERMARREC. *Build One, Get One Free: Leveraging the Coexistence of Multiple P2P Overlay Networks*, in "Proceedings of ICDCS 2007", Toronto, Canada, June 2007

[10] A. MOSTEFAOUI, S. RAJSBAUM, M. RAYNAL, C. TRAVERS. *From Diamond W to Omega: a simple bounded quiescent reliable broadcast-based transformation*, in "Journal of Parallel and Distributed Computing", 2007, vol. 61, n^{o} 1, pp. 125–129

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. BOUTET. , *Décentralisation des systèmes de personnalisation*, Université Rennes 1, March 2013, http://hal.inria.fr/tel-00861370

[12] T. CRAIN. , *Faciliter l'utilisation des mémoires transactionnelles logicielles*, Université Rennes 1, March 2013, http://hal.inria.fr/tel-00861274

[13] K. KLOUDAS. , *Exploitation du contenu pour l'optimisation du stockage distribué*, Université Rennes 1, March 2013, http://hal.inria.fr/tel-00806078

[14] A. VAN KEMPEN. , *Optimiser l'utilisation de la bande passante dans les systèmes de stockage distribué*, Université Rennes 1, March 2013, http://hal.inria.fr/tel-00862845

Articles in International Peer-Reviewed Journals

[15] X. BAI, R. GUERRAOUI, A.-M. KERMARREC. *Personalizing Top-k Processing On-line in a Peer-to-Peer Social Tagging Network*, in "ACM Transactions on Internet Technology", 2014, http://hal.inria.fr/hal-00925990

[16] M. BERTIER, M. OBROVAC, C. TEDESCHI. *Adaptive atomic capture of multiple molecules*, in "Journal of Parallel and Distributed Computing", September 2013, vol. 73, n^{o} 9, pp. 1251-1266, http://hal.inria.fr/hal-00915220

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

[18] A. BOUTET, E. YONEKI, K. HYOUNGSHICK. *What's in Twitter I Know What Parties are Popular and Who You are Supporting Now!*, in "Social Network Analysis and Mining", July 2013, http://hal.inria.fr/hal-00849890

[19] C. DELPORTE-GALLET, H. FAUCONNIER, R. GUERRAOUI, A.-M. KERMARREC, E. RUPPERT, H. TRAN-THE. *Byzantine agreement with homonyms*, in "Distributed Computing", 2013, vol. 26, n^{o} 5-6, pp. 321-340 [*DOI :* 10.1007/S00446-013-0190-3], http://hal.inria.fr/hal-00839625

[20] D. FREY, A. JÉGOU, A.-M. KERMARREC, M. RAYNAL, J. STAINER. *Trust-Aware Peer Sampling: Performance and Privacy Tradeoffs*, in "Journal of Theoretical Computer Science (TCS)", February 2013, http://hal.inria.fr/hal-00872996

[21] A. GIURGIU, R. GUERRAOUI, K. HUGUENIN, A.-M. KERMARREC. *Computing in Social Networks*, in "Journal of Information and Computation", 2013, pp. 1-14, SSS 2010 Special Issue [*DOI :* 10.1016/J.IC.2013.11.001], http://hal.inria.fr/hal-00827745

[22] A.-M. KERMARREC, P. TRIANTAFILLOU. *XL Peer-to-Peer Pub/Sub Systems*, in "ACM Computing Surveys", 2014, vol. 46, n^{o} 2, http://hal.inria.fr/hal-00853828

Invited Conferences

[23] F. TAÏANI. *Deconstructing Complex Distributed Platforms: A Report From the Trenches*, in "9th International Workshop on Automated Specification and Verification of Web Systems", Florence, Italy, A. RAVARA, J. SILVA (editors), Electronic Proceedings in Theoretical Computer Science, June 2013, vol. 123, 2 p. [*DOI :* 10.4204/EPTCS.123], http://hal.inria.fr/hal-00919426

International Conferences with Proceedings

[24] D. ALISTARH, J. ASPNES, G. GIAKKOUPIS, P. WOELFEL. *Randomized loose renaming in O(log log n) time*, in "32nd ACM Symposium on Principles of Distributed Computing (PODC)", Montreal, Canada, July 2013, http://hal.inria.fr/hal-00856744

[25] N. AMANN, A. GOUTA, D. HONG, A.-M. KERMARREC, Y. LE LOUEDEC. *Large scale analysis of HTTP Adaptive Streaming over the Mobile Networks*, in "15èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel)", Pornic, France, N. NISSE, F. ROUSSEAU, Y. BUSNEL (editors), April 2013, pp. 1-4, http://hal.inria.fr/hal-00813595

[26] X. BAI, A. JÉGOU, F. JUNQUEIRA, V. LEROY. *DynaSoRe: Efficient In-Memory Store for Social Applications*, in "Middleware", Beijing, China, December 2013, http://hal.inria.fr/hal-00872990

[27] X. BAI, A. JÉGOU, F. JUNQUEIRA, V. LEROY. *DynaSoRe: Efficient In-Memory Store for Social Applications*, in "ACM/IFIP/USENIX 14th International Middleware Conference", Beijing, China, December 2013, pp. 425-444 [*DOI :* 10.1007/978-3-642-45065-5_22], http://hal.inria.fr/hal-00932468

[28] J. BOURGEOIS, J. CAO, M. RAYNAL, D. DHOUTAUT, J. PIRANDA, E. DEDU, A. MOSTEFAOUI, H. MABED. *Coordination and Computation in distributed intelligent MEMS*, in "AINA 2013, 27th IEEE Int. Conf. on Advanced Information Networking and Applications", Spain, March 2013, pp. 118–123, http://hal.inria.fr/hal-00931533

[29] A. BOUTET, D. FREY, R. GUERRAOUI, A. JÉGOU, A.-M. KERMARREC. *WhatsUp Decentralized Instant News Recommender*, in "IPDPS 2013", Boston, United States, May 2013, http://hal.inria.fr/hal-00769291

[30] A. BOUTET, D. FREY, A. JÉGOU, A.-M. KERMARREC, H. BORBA RIBEIRO. *FreeRec: an Anonymous and Distributed Personalization Architecture*, in "NETYS", Marrakesh, Morocco, 2013, http://hal.inria.fr/hal-00820377

[31] A. BOUTET, K. KLOUDAS, A.-M. KERMARREC. *FStream: a decentralized and social music streamer*, in "NETYS", Marrakech, Morocco, May 2013, http://hal.inria.fr/hal-00828542

[32] G. GIAKKOUPIS, M. HELMI, L. HIGHAM, P. WOELFEL. *An O(sqrt(n)) space bound for obstruction-free leader election*, in "DISC - 27th International Symposium on Distributed Computing", Jerusalem, Israel, October 2013, http://hal.inria.fr/hal-00875167

[33] G. GIAKKOUPIS, A.-M. KERMARREC, P. WOELFEL. *Gossip protocols for renaming and sorting*, in "DISC - 27th International Symposium on Distributed Computing", Jerusalem, Israel, October 2013, http://hal.inria.fr/hal-00875162

[34] A. GOUTA, C. HONG, D. HONG, A.-M. KERMARREC, Y. LELOUEDEC. *Large scale analysis of HTTP adaptive streaming in mobile networks*, in "Proc. of the IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'13)", Spain, August 2013, 10 p. , http://hal.inria.fr/hal-00833496

[35] A. GOUTA, D. HONG, A.-M. KERMARREC, Y. LELOUEDEC. *HTTP adaptive streaming in mobile networks: characteristics and caching opportunities*, in "IEEE 21st International Symposium on Modeling, Analysis and

Simulation of Computer and Telecommunication Systems", San Francisco, United States, June 2013, 10 p. , http://hal.inria.fr/hal-00833493

[36] R. GUERRAOUI, F. HUC, A.-M. KERMARREC. *Highly Dynamic Distributed Computing with Byzantine Failures*, in "ACM Symposium on Principles of Distributed Computing", Montreal, Canada, ACM, 2013, http://hal.inria.fr/hal-00870960

[37] A.-M. KERMARREC, A. MOIN. *FlexGD : A Flexible Force-directed Model for Graph Drawing*, in "IEEE PacificVis", Sydney, Australia, 2013, http://hal.inria.fr/hal-00764245

[38] M. RAYNAL, J. STAINER. *Simultaneous Consensus vs Set Agreement: A Message-Passing-Sensitive Hierarchy of Agreement Problems.*, in "SIROCCO", Ischia, Italy, Lecture Notes in Computer Science, Thomas Moscibroda and Adele A. Rescigno, July 2013, vol. 8179, pp. 298-309 [*DOI :* 10.1007/978-3-319-03578-9_25], http://hal.inria.fr/hal-00920725

[39] M. RAYNAL, J. STAINER. *Synchrony weakened by message adversaries vs asynchrony restricted by failure detectors*, in "PODC", Montréal, Canada, ACM, July 2013, pp. 166-175 [*DOI :* 10.1145/2484239.2484249], http://hal.inria.fr/hal-00920734

[40] F. TAÏANI, B. PORTER, G. COULSON, M. RAYNAL. *Cliff-Edge Consensus: Agreeing on the Precipice*, in "12th International Conference on Parallel Computing Technologies (PaCT-2013)", St. Petersburg, Russian Federation, V. MALYSHKIN (editor), Lecture Notes in Computer Science, Springer, September 2013, vol. 7979, pp. 51-64 [*DOI :* 10.1007/978-3-642-39958-9_5], http://hal.inria.fr/hal-00876054

Scientific Books (or Scientific Book chapters)

[41] M. RAYNAL. , *Concurrent Programming: Algorithms, Principles, and Foundations*, Springer, 2013, XXXII + 515 p. , http://hal.inria.fr/hal-00922217

[42] M. RAYNAL. , *Distributed Algorithms for Message-Passing Systems*, Springer, 2013, XXX + 500 p. , http://hal.inria.fr/hal-00922219

Research Reports

[43] M. BERTIER, M. PERRIN, C. TEDESCHI. , *On the Complexity of Concurrent Multiset Rewriting*, Inria, December 2013, n° RR-8408, 17 p. , http://hal.inria.fr/hal-00912554

[44] A. BOUTET, D. FREY, A. JÉGOU, A.-M. KERMARREC, H. BORBA RIBEIRO. , *FreeRec: an Anonymous and Distributed Personalization Architecture*, July 2013, 20 p. , http://hal.inria.fr/hal-00844813

[45] A. BOUTET, D. FREY, A.-M. KERMARREC, R. GUERRAOUI. , *Democratizing Personalization*, Inria, March 2013, n° RR-8254, http://hal.inria.fr/hal-00799221

[46] A. BOUTET, A.-M. KERMARREC, D. FREY, R. GUERRAOUI, A. JÉGOU. , *Privacy-Preserving Distributed Collaborative Filtering*, Inria, March 2013, n° RR-8253, http://hal.inria.fr/hal-00799209

[47] M. HERLIHY, S. RAJSBAUM, M. RAYNAL, J. STAINER. , *Computing in the Presence of Concurrent Solo Executions*, May 2013, n° PI-2004, http://hal.inria.fr/hal-00825619

[48] A. LÈBRE, J. PASTOR, M. BERTIER, F. DESPREZ, J. ROUZAUD-CORNABAS, C. TEDESCHI, P. ANEDDA, G. ZANETTI, R. NOU, T. CORTES, E. RIVIÈRE, T. ROPARS. , *Beyond The Cloud, How Should Next Generation Utility Computing Infrastructures Be Designed?*, Inria, July 2013, n^o RR-8348, http://hal.inria.fr/hal-00854204

[49] S. RAJSBAUM, R. MICHEL. , *An Introductory Tutorial to Concurrency-Related Distributed Recursion*, June 2013, n^{o} PI 2006, 14 p. , http://hal.inria.fr/hal-00858195

[50] M. RAYNAL, J. STAINER, J. CAO, W. WU. , *A Simple Broadcast Algorithm for Recurrent Dynamic Systems*, September 2013, n^{o} PI-2008, http://hal.inria.fr/hal-00862442

[51] M. RAYNAL, J. STAINER. , *Round-based Synchrony Weakened by Message Adversaries vs Asynchrony Enriched with Failure Detectors*, February 2013, n^{o} PI-2002, http://hal.inria.fr/hal-00787978

[52] M. RAYNAL, J. STAINER. , *Simultaneous Consensus vs Set Agreement a Message-Passing Sensitive Hierarchy of Agreement Problems*, February 2013, n^{o} PI-2003, http://hal.inria.fr/hal-00787992

[53] C. TRAVERS, S. RAJSBAUM, M. RAYNAL. , *The Iterated Restricted Immediate Snapshot Model*, June 2013, n^{o} PI-2005, http://hal.inria.fr/hal-00829436

Other Publications

[54] A. CASTANEDA, D. IMBS, S. RAJSBAUM, R. MICHEL. , *Generalized Symmetry Breaking Tasks*, September 2013, http://hal.inria.fr/hal-00862230

References in notes

[55] M. AGUILERA. *A Pleasant Stroll Through the Land of Infinitely Many Creatures*, in "ACM SIGACT News, Distributed Computing Column", 2004, vol. 35, n^{o} 2

[56] D. ANGLUIN. *Local and Global Properties in Networks of Processes*, in "Proc. 12th ACM Symposium on Theory of Computing (STOC'80)", 1980

[57] K. BIRMAN, M. HAYDEN, O. OZKASAP, Z. XIAO, M. BUDIU, Y. MINSKY. *Bimodal Multicast*, in "ACM Transactions on Computer Systems", May 1999, vol. 17, n^{o} 2, pp. 41-88

[58] A. DEMERS, D. GREENE, C. HAUSER, W. IRISH, J. LARSON, S. SHENKER, H. STURGIS, D. SWINEHART, D. TERRY. *Epidemic algorithms for replicated database maintenance*, in "Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC'87)", August 1987

[59] P. EUGSTER, S. HANDURUKANDE, R. GUERRAOUI, A.-M. KERMARREC, P. KOUZNETSOV. *Lightweight Probabilistic Broadcast*, in "ACM Transaction on Computer Systems", November 2003, vol. 21, n^{o} 4

[60] L. LAMPORT. *Time, clocks, and the ordering of events in distributed systems*, in "Communications of the ACM", 1978, vol. 21, n^{o} 7

[61] M. MERRITT, G. TAUBENFELD. *Computing Using Infinitely Many Processes*, in "Proc. 14th Int'l Symposium on Distributed Computing (DISC'00)", 2000

[62] S. RATNASAMY, P. FRANCIS, M. HANDLEY, R. KARP, S. SHENKER. *A Scalable Content-Addressable Network*, in "Conference of the Special Interest Group on Data Communication (SIGCOMM'01)", 2001

[63] A. ROWSTRON, P. DRUSCHEL. *Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems*, in "IFIP/ACM Intl. Conf. on Distributed Systems Platforms (Middleware)", 2001

[64] I. STOICA, R. MORRIS, D. KARGER, F. KAASHOEK, H. BALAKRISHNAN. *Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications*, in "SIGCOMM'01", 2001

[65] S. VOULGARIS, D. GAVIDIA, M. VAN STEEN. *CYCLON: Inexpensive Membership Management for Unstructured P2P Overlays*, in "Journal of Network and Systems Management", 2005, vol. 13, n^{o} 2