IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Rennes

Université Rennes 1

Activity Report 2012

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 Services

Table of contents

Project-Team ASAP

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

Creation of the Project-Team: July 01, 2007 .

1. Members

Research Scientists

Anne-Marie Kermarrec [Team Leader, Research Director, INRIA, HdR]
Davide Frey [Junior Researcher, INRIA]
George Giakkoupis [Junior Researcher, INRIA, (since March 2012)]

Faculty Members

Michel Raynal [Professor (Pr), University Rennes 1, HdR]
Marin Bertier [Assistant Professor (MdC), INSA Rennes]
François Taïani [Professor (Pr), University Rennes 1 (since November 2012), HdR]
Stéphane Weiss [ATER until August 2012]

Engineer

Heverson Borba Ribeiro [Ingénieur-Expert INRIA]

PhD Students

Mohammad Alaggan [MENRT Grant]
Antoine Boutet [INRIA Grant]
Tyler Crain [Marie-Curie European Grant]
Ali Gouta [Cifre Orange Labs Grant]
Damien Imbs [MENRT Grant until April 2012]
Arnaud Jegou [INRIA Grant]
Eleni Kanellou [Marie-Curie European Grant]
Konstantinos Kloudas [INRIA Grant]
Afshin Moin [INRIA Grant until July 2012]
Antoine Rault [INRIA & Regional Grant]
Julien Stainer [MESR Grant]
Alexandre Van Kempen [Cifre Technicolor Grant]

Post-Doctoral Fellow

Armando Castaneda [until May 2012]

Administrative Assistant

Cécile Bouton [INRIA]

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 thousand 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 algorithm 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.

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 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 protocol 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 rumour 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

- **Best Paper Award PODC 2012** ACM Symposium on Principles of Distributed Computing (G. Giakkoupis and P. Woelfel, *On the time and space complexity of randomized test-and-set*).
- **Chair of the ACM Software System Award committee** (A.-M. Kermarrec)
- **ERC Proof of Concept Grant** (A.-M. Kermarrec)

3. Scientific Foundations

3.1. Distributed Computing

Distributed computing 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*" [56] (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 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 [57], [50], [51]. 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 thousand 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 [58], [59], [60] following a specific topology or an unstructured network [55], [61] 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 [53], [52], 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 the 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 [52], 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 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

5.1. WhatsUp: A Distributed News Recommender

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

Contact:	Antoine Boutet
Licence:	Open Source
Presentation:	A Distributed News Recommender
Status:	Beta version

This work has lead to the development of WhatsUp, a distributed recommendation system aimed to distribute instant news in a large scale dynamic system. WhatsUp has two parts, an embedded application server in order to exchange with other peers in the system and a fully dynamic web interface for displaying news and collecting opinions about what the user reads. Underlying this web-based application lies Beep, a biased epidemic dissemination protocol that delivers news to interested users in a fast manner while limiting spam. Beep is parametrized on the fly to manage the orientation and the amplification of news dissemination. Every user forwards the news of interest to a randomly selected set of users with a preference towards those that have similar interests (orientation). The notion of interest does not rely on any explicit social network or subscription scheme, but rather on an implicit and dynamic overlay capturing the commonalities between users with respect to what they are interested in. The size of the set of users to which a news is forwarded depends on the interest of the news (amplification). A centralized version of WhatsUp is already up and running and the decentralized one is still in beta version.

5.2. GossipLib: effective development of gossip-based applications

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

Contact:	Davide Frey
Licence:	Open Source
Presentation:	Library for Gossip protocols
Status:	released version 0.7alpha

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 applications: it thus provides 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 simply by changing one line in a configuration file.

5.3. YALPS

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

Contact:	Davide Frey
Licence:	Open Source
Presentation:	Library for Gossip protocols
Status:	released version 0.3alpha

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 without a single change in the code. 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.

The work has been done in collaboration with Maxime Monod (EPFL).

5.4. HEAP: Heterogeneity-aware gossip protocol.

Participants: Davide Frey, Arnaud Jegou, 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

This section summarizes the major results obtained by the ASAP team that relate to the foundations of distributed systems.

6.1.1. Efficient shared memory consensus

Participants: Michel Raynal, Julien Stainer.

This work is on an efficient algorithm that builds a consensus object. This algorithm is based on an Ω failure detector (to obtain consensus liveness) and a store-collect object (to maintain its safety). A store-collect object provides the processes with two operations, a store operation which allows the invoking process to deposit a new value while discarding the previous value it has deposited and a collect operation that returns to the invoking process a set of pairs (i, val) where val is the last value deposited by the process p_i. A store-collect object has no sequential specification.

While store-collect objects have been used as base objects to design wait-free constructions of more sophisticated objects (such as snapshot or renaming objects), as far as we know, they have not been explicitly used to built consensus objects. The proposed store-collect-based algorithm, which is round-based, has several noteworthy features. First it uses a single store-collect object (and not an object per round). Second, during a round, a process invokes at most once the store operation and the value val it deposits is a simple pair $\langle r, v\rangle$ where r is a round number and v a proposed value. Third, a process is directed to skip rounds according to its view of the current global state (thereby saving useless computation rounds). Finally, the proposed algorithm benefits from the adaptive wait-free implementations that have been proposed for store-collect objects, namely, the number of shared memory accesses involved in a collect operation is $O(k)$ where k is the number of processes that have invoked the store operation. This makes this new algorithm particularly efficient and interesting for multiprocess programs made up of asynchronous crash-prone processes that run on top of multicore architectures.

6.1.2. A Contention-Friendly, Non-blocking Skip List

Participants: Tyler Crain, Michel Raynal.

This work [27] presents a new non-blocking skip list algorithm. The algorithm alleviates contention by localizing synchronization at the least contended part of the structure without altering consistency of the implemented abstraction. The key idea lies in decoupling a modification to the structure into two stages: an eager abstract modification that returns quickly and whose update affects only the bottom of the structure, and a lazy selective adaptation updating potentially the entire structure but executed continuously in the background. As non-blocking skip lists are becoming appealing alternatives to latch-based trees in modern main-memory databases, we integrated it into a main-memory database benchmark, SPECjbb. On SPECjbb as well as on micro-benchmarks, we compared the performance of our new non-blocking skip list against the performance of the JDK non-blocking skip list. Results indicate that our implementation is up to 2:5 faster than the JDK skip list.

6.1.3. STM Systems: Enforcing Strong Isolation between Transactions and Non-transactional Code

Participants: Tyler Crain, Eleni Kanellou, Michel Raynal.

Transactional memory (TM) systems implement the concept of an atomic execution unit called a transaction in order to discharge programmers from explicit synchronization management. But when shared data is atomically accessed by both transaction and non-transactional code, a TM system must provide strong isolation in order to overcome consistency problems. Strong isolation enforces ordering between non-transactional operations and transactions and preserves the atomicity of a transaction even with respect to non-transactional code. This work [29] presents a TM algorithm that implements strong isolation with the following features: (a) concurrency control of non-transactional operations is not based on locks and is particularly efficient, and (b) any non-transactional read or write operation always terminates (there is no notion of commit/abort associated with them).

6.1.4. A speculation-friendly binary search tree

Participants: Tyler Crain, Michel Raynal.

In this work [26], in collaboration with Vincent Gramoli, we introduce the first binary search tree algorithm designed for speculative executions. Prior to this work, tree structures were mainly designed for their pessimistic (non-speculative) accesses to have a bounded complexity. Researchers tried to evaluate transactional memory using such tree structures whose prominent example is the red-black tree library developed by Oracle Labs that is part of multiple benchmark distributions. Although well-engineered, such structures remain badly suited for speculative accesses, whose step complexity might raise dramatically with contention. We show that our speculation-friendly tree outperforms the existing transaction-based version of the AVL and the red-black trees. Its key novelty stems from the decoupling of update operations: they are split into one transaction that modifies the abstraction state and multiple ones that restructure its tree implementation in the background. In particular, the speculation-friendly tree is shown correct, reusable and it speeds up a transaction-based travel reservation application by up to 3:5.

6.1.5. Towards a universal construction for transaction-based multiprocess programs

Participants: Tyler Crain, Damien Imbs, Michel Raynal.

The aim of a Software Transactional Memory (STM) system is to discharge the programmer from the explicit management of synchronization issues. The programmer's job resides in the design of multiprocess programs in which processes are made up of transactions, each transaction being an atomic execution unit that accesses concurrent objects. The important point is that the programmer has to focus her/his efforts only on the parts of code which have to be atomic execution units without worrying on the way the corresponding synchronization has to be realized. Non-trivial STM systems allow transactions to execute concurrently and rely on the notion of commit/abort of a transaction in order to solve their conflicts on the objects they access simultaneously. In some cases, the management of aborted transactions is left to the programmer. In other cases, the underlying system scheduler is appropriately modified or an underlying contention manager is used in order that each transaction be ("practically always" or with high probability) eventually committed. This work [28] presents a deterministic STM system in which (1) every invocation of a transaction is executed exactly once and (2) the notion of commit/abort of a transaction remains unknown to the programmer. This system, which imposes restriction neither on the design of processes nor or their concurrency pattern, can be seen as a step in the design of a deterministic universal construction to execute transaction-based multiprocess programs on top of a multiprocessor. Interestingly, the proposed construction is lock-free (in the sense that it uses no lock).

6.1.6. A Tight RMR Lower Bound for Randomized Mutual Exclusion

Participant: George Giakkoupis.

The Cache Coherent (CC) and the Distributed Shared Memory (DSM) models are standard shared memory models, and the Remote Memory Reference (RMR) complexity is considered to accurately predict the actual performance of mutual exclusion algorithms in shared memory systems. Through a collaboration with Philipp Woelfel [32], we proved a tight lower bound for the RMR complexity of deadlock-free randomized mutual exclusion algorithms in both the CC and the DSM model with atomic registers and compare&swap objects and an adaptive adversary. Our lower bound establishes that an adaptive adversary can schedule n processes in such a way that each enters the critical section once, and the total number of RMRs is $\Omega(n \log n / \log \log n)$ in expectation. This matches an upper bound of Hendler and Woelfel (2011).

6.1.7. On the Time and Space Complexity of Randomized Test-And-Set

Participant: George Giakkoupis.

Through a collaboration with Philipp Woelfel [33] we studied the time and space complexity of randomized Test-And-Set (TAS) implementations from atomic read/write registers in asynchronous shared memory models with n processes. We presented an adaptive TAS implementation with an expected (individual) step complexity of $O(\log^* k)$, for contention k, against the oblivious adversary, improving a previous (non-adaptive) upper bound of $O(\log \log n)$ by Alistarh and Aspnes (2011). We also showed how to modify the adaptive RatRace TAS algorithm by Alistarh, Attiya, Gilbert, Giurgiu, and Guerraoui (2010) to improve the space complexity from $O(n^3)$ to $O(n)$, while maintaining logarithmic expected step complexity against the adaptive adversary. Finally, we proved that for any randomized 2-process TAS algorithm there exists a schedule determined by an oblivious adversary, such that with probability at least $1/4^t$ one of the processes does not finish its TAS operation in within fewer than t steps. This complements a lower bound by Attiya and Censor-Hillel (2010) of a similar result for $n \geq 3$ processes.

6.2. Large-scale and user-centric distributed system

6.2.1. WhatsUp: P2P news recommender

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

The main application in the context of GOSSPLE is WhatsUp, an instant news system designed for a large-scale network with no central authority. WhatsUp builds an implicit social network based on the opinions users express about the news items they receive (like-dislike). This is achieved through an obfuscation mechanism that does not require users to ever reveal their exact profiles. WhatsUp disseminates news items through a novel heterogeneous gossip protocol that biases the choice of its targets towards those with similar interests and amplifies dissemination based on the level of interest in every news item. WhatsUp outperforms various alternatives in terms of accurate and complete delivery of relevant news items while preserving the fundamental advantages of standard gossip: namely simplicity of deployment and robustness. This work has been carried out in collaboration with Rachid Guerraoui from EPFL and was demonstrated during the different local events and will appear in IPDPS 2013 [21].

6.2.2. *Privacy in P2P recommenders*

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

We also propose a mechanism to preserve privacy in WhatsUp, which can also be used in any distributed recommendation system. Our approach relies on (i) an original obfuscation mechanism hiding the exact profiles of users without significantly decreasing their utility, as well as (ii) a randomized dissemination algorithm ensuring differential privacy during the dissemination process. Results show that our solution preserves accuracy without the need for users to reveal their preferences. Our approach is also flexible and robust to censorship.

6.2.3. *BLIP: Non-interactive differentially-private similarity computation on Bloom filters*

Participants: Mohammad Alaggan, Anne-Marie Kermarrec.

In this project [19], done in collaboration with Sébastien Gambs (team CIDRE), we consider the scenario in which the profile of a user is represented in a compact way, as a Bloom filter, and the main objective is to privately compute in a distributed manner the similarity between users by relying only on the Bloom filter representation. In particular, we aim at providing a high level of privacy with respect to the profile even if a potentially unbounded number of similarity computations take place, thus calling for a non-interactive mechanism. To achieve this, we propose a novel non-interactive differentially private mechanism called BLIP (for BLoom-and-flIP) for randomizing Bloom filters. This approach relies on a bit flipping mechanism and offers high privacy guarantees while maintaining a small communication cost. Another advantage of this non-interactive mechanism is that similarity computation can take place even when the user is offline, which is impossible to achieve with interactive mechanisms. Another of our contributions is the definition of a probabilistic inference attack, called the "Profile Reconstruction Attack", that can be used to reconstruct the profile of an individual from his Bloom filter representation, along with the "Profile Distinguishing Game". More specifically, we provide an analysis of the protection offered by BLIP against this profile reconstruction attack by deriving an upper and lower bound for the required value of the differential privacy parameter ϵ.

6.2.4. *Heterogeneous Differential Privacy*

Participants: Mohammad Alaggan, Anne-Marie Kermarrec.

The massive collection of personal data by personalization systems has rendered the preservation of privacy of individuals more and more difficult. Most of the proposed approaches to preserve privacy in personalization systems usually address this issue uniformly across users, thus completely ignoring the fact that users have different privacy attitudes and expectations (even among their own personal data). In this project, in collaboration with Sébastien Gambs (team CIDRE), we propose to account for this non-uniformity of privacy expectations by introducing the concept of heterogeneous differential privacy. This notion captures both the variation of privacy expectations among users as well as across different pieces of information related to the same user. We also describe an explicit mechanism achieving heterogeneous differential privacy, which is a modification of the Laplacian mechanism due to Dwork [54], we evaluate on real datasets the impact of the proposed mechanism with respect to a semantic clustering task. The results of our experiments clearly demonstrate that heterogeneous differential privacy can account for different privacy attitudes while sustaining a good level of utility as measured by the recall.

6.2.5. Social Market

Participants: Davide Frey, Arnaud Jegou, 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) solve 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 a-la 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. We submitted this work for journal publication.

6.2.6. Geolocated Social Networks

Participants: Anne-Marie Kermarrec, François Taïani.

Geolocated social networks, that combine traditional social networking features with geolocation information, have grown tremendously over the last few years. Yet, very few works have looked at implementing geolocated social networks in a fully distributed manner, a promising avenue to handle the growing scalability challenges of these systems. In [25], we have focused on georecommendation, and showed that existing decentralized recommendation mechanisms perform in fact poorly on geodata. In this work, we have proposed a set of novel gossip-based mechanisms to address this problem, and captured these mechanisms in a modular similarity framework called "Geology". The resulting platform is lightweight, efficient, and scalable. More precisely, we have shown its benefits in terms of recommendation quality and communication overhead on a real data set of 15,694 users from Foursquare, a leading geolocated social network.

6.2.7. Content and Geographical Locality in User-Generated Content Sharing Systems

Participants: Anne-Marie Kermarrec, Konstantinos Kloudas, François Taïani.

User Generated Content (UGC), such as YouTube videos, accounts for a substantial fraction of the Internet traffic. To optimize their performance, UGC services usually rely on both proactive and reactive approaches that exploit spatial and temporal locality in access patterns. Alternative types of locality are also relevant and hardly ever considered together. In [34], we show on a large (more than $650,000$ videos) YouTube dataset that content locality (induced by the related videos feature) and geographic locality, are in fact correlated. More specifically, we show how the geographic view distribution of a video can be inferred to a large extent from that of its related videos. We leverage these findings to propose a UGC storage system that proactively places videos close to the expected requests. Compared to a caching-based solution, our system decreases by 16% the number of requests served from a different country than that of the requesting user, and even in this case, the distance between the user and the server is 29% shorter on average.

6.2.8. Probabilistic Deduplication for Cluster-Based Storage Systems

Participants: Davide Frey, Anne-Marie Kermarrec, Konstantinos Kloudas.

The need to backup huge quantities of data has led to the development of a number of distributed deduplication techniques that aim to reproduce the operation of centralized, single-node backup systems in a cluster-based environment. At one extreme, stateful solutions rely on indexing mechanisms to maximize deduplication. However the cost of these strategies in terms of computation and memory resources makes them unsuitable for large-scale storage systems. At the other extreme, stateless strategies store data blocks based only on their content, without taking into account previous placement decisions, thus reducing the cost but also the effectiveness of deduplication. In [30], we propose, Produck, a stateful, yet lightweight cluster-based backup system that provides deduplication rates close to those of a single-node system at a very low computational cost and with minimal memory overhead. In doing so, we provide two main contributions: a lightweight probabilistic node-assignment mechanism and a new bucketbased load-balancing strategy. The former allows Produck to quickly identify the servers that can provide the highest deduplication rates for a given data block. The latter efficiently spreads the load equally among the nodes. Our experiments compare Produck against state-of-the-art alternatives over a publicly available dataset consisting of 16 full *Wikipedia* backups, as well as over a private one consisting of images of the environments available for deployment on the Grid5000 experimental platform. Our results show that, on average, Produck provides (i) up to 18% better deduplication compared to a stateless minhash-based technique, and (ii) an 18-fold reduction in computational cost with respect to a stateful BloomFilter-based solution.

6.2.9. Large scale analysis of HTTP adaptive streaming in mobile networks

Participants: Ali Gouta, Anne-Marie Kermarrec.

In collaboration with Yannick Le Louedec and Nathalie Amann we have been working in the context of adaptive streaming in mobile networks. HTTP Adaptive bitrate video Streaming (HAS) is now widely adopted by Content Delivery Network Providers (CDNPs) and Telecom Operators (Telcos) to improve user Quality of Experience (QoE). In HAS, several versions of videos are made available in the network so that the quality of the video can be chosen to better fit the bandwidth capacity of users. These delivery requirements raise new challenges with respect to content caching strategies, since several versions of the content may compete to be cached. We used a real HAS dataset collected in France and provided by a mobile telecom operator involving more than 485,000 users requesting adaptive video contents through more than 8 million video sessions over a 6 week measurement period. Firstly, we proposed a fine-grained definition of content popularity by exploiting the segmented nature of video streams. We also provided analysis about the behavior of clients when requesting such HAS streams. We proposed novel caching policies tailored for chunk-based streaming. Then we studied the relationship between the requested video bitrates and radio constraints. Finally, we studied the users' patterns when selecting different bitrates of the same video content. Our findings provide useful insights that can be leveraged by the main actors of video content distribution to improve their content caching strategy for adaptive streaming contents as well as to model users' behavior in this context.

6.2.10. Regenerating Codes: A System Perspective

Participants: Anne-Marie Kermarrec, Alexandre van Kempen.

The explosion of the amount of data stored in cloud systems calls for more efficient paradigms for redundancy. While replication is widely used to ensure data availability, erasure correcting codes provide a much better trade-off between storage and availability. Regenerating codes are good candidates for they also offer low repair costs in term of network bandwidth. While they have been proven optimal, they are difficult to understand and parameterize. In collaboration with Nicolas Le Scouarnec, Gilles Straub and Steve Jiekak from Technicolor, we performed an analysis of regenerating codes, which enables practitioners to grasp the various trade-offs. More specifically we made two contributions: (i) we studied the impact of the parameters by conducting an analysis at the level of the system, rather than at the level of a single device; (ii) we compared the computational costs of various implementations of codes and highlight the most efficient ones. Our goal is to provide system designers with concrete information to help them choose the best parameters and design for regenerating codes.

6.2.11. Availability-based methods for distributed storage systems

Participants: Anne-Marie Kermarrec, Alexandre van Kempen.

Distributed storage systems rely heavily on redundancy to ensure data availability as well as durability. In networked systems subject to intermittent node unavailability, the level of redundancy introduced in the system should be minimized and maintained upon failures. Repairs are well- known to be extremely bandwidth-consuming and it has been shown that, without care, they may significantly congest the system. In collaboration with Gilles Straub and Erwan Le Merrer from Technicolor, we proposed an approach to redundancy management accounting for nodes heterogeneity with respect to availability. We show that by using the availability history of nodes, the performance of two important faces of distributed storage (replica placement and repair) can be significantly improved. Replica placement is achieved based on complementary nodes with respect to nodes availability, improving the overall data availability. Repairs can be scheduled thanks to an adaptive per-node timeout according to node availability, so as to decrease the number of repairs while reaching comparable availability. We propose practical heuristics for those two issues. We evaluate our approach through extensive simulations based on real and well-known availability traces. Results clearly show the benefits of our approach with regards to the critical trade-off between data availability, load-balancing and bandwidth consumption.

6.2.12. On The Impact of Users Availability In OSNs

Participants: Antoine Boutet, Anne-Marie Kermarrec, Alexandre van Kempen.

Availability of computing resources has been extensively studied in literature with respect to uptime, session lengths and inter-arrival times of hardware devices or software applications. Interestingly enough, information related to the presence of users in online applications has attracted less attention. Consequently, only a few attempts have been made to leverage user availability pattern to improve such applications. In collaboration with Erwan Le Merrer from Technicolor, we studied an availability trace collected from MySpace. Our results show that the online presence of users tends to be correlated to that of their friends. User availability also plays an important role in some algorithms and focus on information spreading. In fact, identifying central users i.e. those located in central positions in a network, is key to achieve a fast dissemination and the importance of users in a social graph precisely vary depending on their availability.

6.2.13. Chemical programming model

Participant: Marin Bertier.

This work, done in collaboration with the Myriads project team, focuses on chemical programming, a promising paradigm to design autonomic systems. The metaphor envisions a computation as a set of concurrent reactions between molecules of data arising non-deterministically, until no more reactions can take place, in which case, the solution contains the final outcome of the computation.

More formally, such models strongly rely on concurrent multiset rewriting: the data are a multiset of molecules, and reactions are the application of a set of conditioned rewrite rules. At run time, these rewritings are applied concurrently, until no rule can be applied anymore (the elements they need do not exist anymore in the multiset). One of the main barriers towards the actual adoption of such models come from their complexity at run time: each computation step may require a complexity in $O(n^k)$ where n denotes the number of elements in the multiset, and k the size of the subset of elements needed to trigger one rule.

Our objective is to design a distributed chemical platform implementing such concepts. This platform should be adapted to large scale distributed system to benefit at his best the inherent distribution of chemical program.

Within this context, we proposed a protocol for the atomic capture of objects in a DHT [20]. This protocol is distributed and evolving over a large scale platform. As the density of potential request has a significant impact on the liveness and efficiency of such a capture, the protocol proposed is made up of two sub-protocols, each of them aimed at addressing different levels of densities of potential reactions in the solution. While the decision to choose one or the other is local to each node participating in a program's execution, a global coherent behavior is obtained.

7. Bilateral Contracts and Grants with Industry

7.1. Technicolor

Participants: Anne-Marie Kermarrec, Alexandre Van Kempen.

Since 2010, we have had a contract with Technicolor for collaboration on peer-assisted approaches for reliable storage. In this context, Anne-Marie Kermarrec has been the PhD advisor of Alexandre van Kempen since 2010.

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.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. LABEX CominLabs

Participants: Anne-Marie Kermarrec, Davide Frey, Stéphane Weiss.

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.

8.1.2. ANR ARPÈGE project Streams

Participants: Marin Bertier, Michel Raynal, Stéphane Weiss.

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 VERSO project Shaman

Participants: Marin Bertier, Anne-Marie Kermarrec, Michel Raynal.

The Shaman project started in 2009, gathering several members of the team working on distributed systems and distributed algorithms. The aim of this project is to propose new theoretical models for distributed algorithms inspired from real platform characteristics. From these models, we elaborate new algorithms and try to evaluate their theoretical power.

8.1.4. 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) 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 (after 3.5 years of the GOSSPLE ERC SG) 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. AllYours is investigating how to turn such inventions into a successful innovation with high potential targeting both end users and SMEs with an enterprise, semi-centralized, version of the system.

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 will focus 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

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 1 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.

Foundation for Research and Technology Hellas ICS FORTH Greece
Transform

Lancaster University
collaboration on the ERC SG GOSSPLE

Imperial College London
collaboration on the Map-Reduce systems

8.3. International Initiatives

8.3.1. Inria International Partners

University of Calgary

Universidad Nacional Autonoma de Mexico

8.3.2. Participation In International Programs

8.3.2.1. 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

The team welcomed the following research visitors in 2012.

Swan Dubois, Lip 6, 27 January 2012.

Paolo Costa, Imperial College London, from 8 to 10 February 2012 and one week in November.

Rachid Guerraoui, several one week visits in 2012.

Gregor Von Bochmann, University of Ottawa, from 12 to 17 March 2012.

Zekri Lougmiri, Faculté de Sciences d'Oran, 23 April to 4 May 2012.

Zhu Weiping, Hong Kong Polytechnic University, from 15 November 2011 until 14 May 2012.

Anna-Kaisa Pietilainen ; Technicolor Paris, 31 May 2012.

Jean-Pierre Lozzi, Lip 6, 1 June 2012.

Vincent Leroy, Université Joseph Fourier de Grenoble, 29 to 31 October 2012.

Bin Xiao, Hong Kong Polytechnic University, 26 December 2012.

8.4.1. Internships

Mathieu GOESSENS; 6 February 2012 to 6 July 2012. “Peer-to-peer content dissemination”. Supervised by Davide Frey and Anne-Marie Kermarrec.

Ilham IKBAL; 1 March 2012 to 15 August 2012. “Integration du routage en oignon (TOR) dans les protocoles epidemiques”. Supervised by Davide Frey.

Imane ALIFDAL; 1 March 2012 to 31 August 2012. “Integration du routage en oignon (TOR) dans les protocoles epidemiques”. Supervised by Davide Frey.

Benjamin Girault; 19 March to 31 August 2012. “Heterogeneous gossip protocols for news recommendation”. Supervised by Anne-Marie Kermarrec.

Asiff Shaik; 3 August 2012 to 2 January 2013. “Understanding offline social networks and its advantages over the online social network ; resolving some challenges in the offline social networks such as privacy, trust, security and scalability.”. Supervised by Anne-Marie Kermarrec.

8.4.2. Visits to International Teams

Anne-Marie Kermarrec has been a part-time (50%) visiting professor at EPFL Lausanne since September 2012.

9. Dissemination

9.1. Scientific Animation

A.-M. Kermarrec was the chair of the ACM Software System Award Committee.

A.-M. Kermarrec is a member of the scientific committee of the Société Informatique de France

A.-M. Kermarrec gave an invited seminar at the University of Lisbon is September 2012

A.-M. Kermarrec is a member of the steering committee of Eurosys.

A.-M. Kermarrec is a member of the steering committte of Middleware

A.-M. Kermarrec is a member of the steering committte of the SNS workshop

A.-M. Kermarrec is a member of the steering committte of the Winter School on hot topics in distributed systems

A.-M. Kermarrec is a member of the IEEE Internet Computing Editorial Board

A.-M. Kermarrec co-organized the Workshop Inria/Technicolor on distributed storage systems, Nov. 2012

A.-M. Kermarrec was an expert for WWTF (Vienna Science and Technology Fund) in 2012.

A.-M. Kermarrec is a member of the Inria scientific board (Bureau du comité des projets) in Rennes

A.-M. Kermarrec served in the program committees for the following conferences:

Eurosys 2012 : Bern, Swizerland, April 2012

Middleware 2012 : *ACM/IFIP/USENIX International Conference on Middleware*, Montreal, Canada, December 2012

DEBS 2012 : *ACM International Conference on Distributed Event-Based Systems*, Berlin, Germany, July 2012.

LADIS 2012 : *Large-Scale Distributed Systems and Middleware*, Madeira, Portugal, July 2012

VLDB 2013 : *International Conference on Very Large Data Bases*, Trento, Italy, August 2013.

SIGMOD 2013 : *ACM International Conference on Data Management*, NYC, USA, June 2013.

ICDCS 2013 : *International Conference on Distributed Computing Systems*, Philadelphia, USA, July 2013.

M. Raynal was program co-chair of ICDCN 2013, 14th International Conference on Distributed Computing and Networking, Mumbai, January 2013.

M. Raynal served in the program committees for the following conferences:

ICDCN 2012 *13th Int'l Conference on Distributed Computing and Networking (ICDCN'12)*, Hong-Kong, Hong Kong, January 3-6, 2012.

WTTM 2012 *4th Workshop on the Theory of Transactional Memory (WTTM'12)*, (satellite workshop of PODC'12), Madeira, 2012.

DISC 2012 *26th Int'l Symposium on Distributed Computing (DISC'12)*, Salvador, Brazil, 2012.

ICDCS 2012 Liaison Co-Chair, *32th IEEE Int'l Conference on Distributed Computing Systems (ICDCS'12)*, Macau, June, 2012.

D. Frey co-chaired the Workshop on Social Networks Systems colocated with Eurosys, Bern, April 2012.

D. Frey was program co-chair of ICDCN 2013, 14th International Conference on Distributed Computing and Networking, Mumbai, January 2013.

D. Frey served in the technical program committee of the following conferences:

Middleware 2012: *ACM/IFIP/USENIX International Conference on Middleware*, Montreal, Canada, December 2012.

P2P 2012: *The 12th IEEE International Conference on Peer-to-Peer Computing* Tarragona, Spain, September, 2012.

SSS 2012: *The 14th International Symposium on Stabilization, Safety, and Security of Distributed Systems* Toronto, Canada, October 2012.

IPDPS 2013: *IEEE International Parallel & Distributed Processing Symposium*, Boston, Massachusetts USA, May 2013.

A. Boutet served in the technical program committee and shadow program committee for the following conferences:

CyberC 2012: *International Conference on Cyber-enabled distributed computing and knowledge discovery*, Sanya, China, October 2012.

Eurosys 2013: *The European Conference on Computer Systems*, Prague, Czech Republic, April 2013.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence (bachelor) courses:

Marin Bertier, Programmation C, 26h, niveau L3, INSA de Rennes, France.

Marin Bertier, Unix, 20h, niveau L3, INSA de Rennes, France.

Marin Bertier, Programmation Scheme, 36h, niveau L1, INSA de Rennes, France.

Master courses:

Anne-Marie Kermarrec, P2P Systems and Applications, 15h, M2, Université of Rennes 1, France.

Anne-Marie Kermarrec, Gossip-based computing, 5 hours, M2, University of San Sebastian, Spain.

Marin Bertier, Algorithmique distribuée, 16h, niveau M2, INSA de Rennes, France.

Marin Bertier, Système d'exploitation, 56h, niveau M1, INSA de Rennes, France.

Marin Bertier, Parallélisme, 20h, niveau M1, INSA de Rennes, France.

Davide Frey, Scalable Distributed Systems, 10 hours, M2, Université de Rennes 1, France.

Mathieu Goessens, Managing students projects, M2 MIAGE, Université of Evry, France, worked with Delphine Lebédel (Mozilla corporation).

Doctoral courses:

Anne-Marie Kermarrec, Large-scale distributed systems, 28h, EPFL, Switzerland

A.-M. Kermarrec, lecture at the METIS Spring School, Tanger, Marocco.

9.2.2. Supervision

PhD: Damien Imbs, "Calculabilité et conditions de progression des objets partagés en présence de défaillances", Universtité de Rennes 1, April, 2012, advised by Michel Raynal

PhD: Afshin Moin, "Les techniques de recommandation et de visualisation pour les données A Une Grande Echelle", Universtité de Rennes 1, July, 2012, advised by Anne-Marie Kermarrec .

PhD in progress: Mohammad Alaggan, "Private similarity computation in user-centric peer-to-peer systems"; October 1 2010; Anne-Marie and Sébastien Gambs; MENRT Grant.

PhD in progress: Antoine Boutet, Inria Grant (ERC) (2009-2012)

PhD in progress: Tyler Crain, Marie Curie European Grant (2010-2013)

PhD in progress: Ali Gouta, CIFRE ORANGE (2012-2015)

PhD in progress: Arnaud Jégou, Inria Grant (ERC) (2010-2013)

PhD in progress: Eleni Kanellou, "Liveness in Transactional Systems"; September 2011; Michel Raynal; Marie Curie European Grant (2010-2013)

PhD in progress: Konstantinos Kloudas, Inria CORDIS Grant (2009-2012)

PhD in progress: Julien Stainer, MENRT Grant (2011-2014)

PhD in progress: Antoine Rault, "Privacy through decentralization"; October 1, 2012; Anne-Marie Kermarrec, Davide Frey; Inria/Region Grant.

PhD in progress: Alexandre van Kempen, CIFRE TECHNICOLOR (2009-2012)

9.2.3. Juries

Anne-Marie Kermarrec was a member of the following PhD Juries.

Joao Leitao, University of Lisbon, Portugal, August 2012

Adrien Friggeri, ENS Lyon, June 2012

Hervé Baumann, University Paris Diderot, September 2012

Sergey Legtchenko in University Pierre et Marie Curie, October 2012

Raphaël Fournier in University Pierre et Marie Curie, December 2012

Qinna Wang, ENS Lyon, August 2012

Anne-Marie Kermarrec was a member of the following HdR Juries.

Arnaud Legout, Inria Sophia, 20/01/2012

François Taïani was a member of the following PhD juries.

Matthew Leeke, Towards the Design of Efficient Error Detection Mechanisms, The University of Warwick (UK), 16/2/2012 (external examiner)

Simon Meyffret, Local And Social Recommendation In Decentralized Architectures, INSA de Lyon / LIRIS (F), 7/12/2012 (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° 8, p. 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] D. FREY, R. GUERRAOUI, A.-M. KERMARREC, M. MONOD, K. BORIS, M. MARTIN, V. QUÉMA. *Heterogeneous Gossip*, in "Middleware 2009", Urbana-Champaign, IL, USA, 2009, http://hal.inria.fr/inria-00436125/en/.

[5] 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° 7, p. 865–875.

[6] 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° 4, p. 267-278.

[7] 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° 5.

[8] 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.

[9] 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° 1, p. 125–129.

[10] J. PATEL, É. RIVIÈRE, I. GUPTA, A.-M. KERMARREC. *Rappel: Exploiting interest and network locality to improve fairness in publish-subscribe systems*, in "Computer Networks", 2009, vol. 53, n° 13, http://hal.inria.fr/inria-00436057/en/.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] D. IMBS. *Calculabilité et conditions de progression des objets partagés en présence de défaillances*, Université Rennes 1, April 2012, http://hal.inria.fr/tel-00722855.

[12] A. MOIN. *Les Techniques De Recommandation Et De Visualisation Pour Les Données A Une Grande Echelle*, Université Rennes 1, July 2012, http://hal.inria.fr/tel-00724121.

Articles in International Peer-Reviewed Journals

[13] R. GUERRAOUI, K. HUGUENIN, A.-M. KERMARREC, M. MONOD, Ý. VIGFÚSSON. *Decentralized Polling with Respectable Participants*, in "Journal of Parallel and Distributed Computing", January 2012, vol. 72, n° 1 [*DOI :* 10.1016/J.JPDC.2011.09.003], http://hal.inria.fr/inria-00629455.

[14] D. IMBS, M. RAYNAL. *Help when needed, but no more: Efficient read/write partial snapshot*, in "Journal of Parallel and Distributed Computing", 2012, vol. 72, n° 1, p. 1-12, http://hal.inria.fr/hal-00646906.

[15] A.-M. KERMARREC. *Towards a personalised Internet: a case for a full decentralisation*, in "Philosophical transactions of the Royal Society A", 2012, http://hal.inria.fr/hal-00723565.

[16] A.-M. KERMARREC, G. TAN. *Greedy Geographic Routing in Large-Scale Sensor Networks: A Minimum Network Decomposition Approach.*, in "IEEE/ACM Transactions on Networking", June 2012, vol. 20, n° 3, p. 864 -877, http://hal.inria.fr/inria-00619038/en.

[17] F. LE FESSANT, A. PAPADIMITRIOU, A. CARNEIRO VIANA, C. SENGUL, E. PALOMAR. *A Sinkhole Resilient Protocol for Wireless Sensor Networks: Performance and Security Analysis*, in "Computer Communications", January 2012, vol. 35, n° 2, http://hal.inria.fr/hal-00653824.

[18] G. TAN, A.-M. KERMARREC. *Greedy Geographic Routing in Large-Scale Sensor Networks: A Minimum Network Decomposition Approach.*, in "IEEE/ACM Transactions on Networking", 2012, vol. 20, n° 3, p. 864-877, http://hal.inria.fr/hal-00764124.

International Conferences with Proceedings

[19] M. ALAGGAN, S. GAMBS, A.-M. KERMARREC. *BLIP: Non-interactive Differentially-Private Similarity Computation on Bloom Filters*, in "14th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS 2012)", Toronto, Canada, October 2012, http://hal.inria.fr/hal-00724829.

[20] M. BERTIER, M. OBROVAC, C. TEDESCHI. *A Protocol for the Atomic Capture of Multiple Molecules at Large Scale*, in "13th International Conference on Distributed Computing and Networking", Hong-Kong, China, Springer, January 2012, http://hal.inria.fr/hal-00644262.

[21] 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.

[22] A. BOUTET, K. HYOUNGSHICK, E. YONEKI. *What's in Your Tweets? I Know Who You Supported in the UK 2010 General Election*, in "The International AAAI Conference on Weblogs and Social Media (ICWSM)", Dublin, Ireland, June 2012, http://hal.inria.fr/hal-00702390.

[23] A. BOUTET, A.-M. KERMARREC, E. LE MERRER, A. VAN KEMPEN. *On The Impact of Users Availability In OSNs*, in "Social Network Systems (SNS 2012)", Bern, Switzerland, April 2012, http://hal.inria.fr/hal-00702399.

[24] A. BOUTET, E. YONEKI, K. HYOUNGSHICK. *What's in Twitter: I Know What Parties are Popular and Who You are Supporting Now!*, in "2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)", Istanbul, Turkey, August 2012, http://hal.inria.fr/hal-00702405.

[25] J. CARRETERO, F. ISAILA, A.-M. KERMARREC, F. TAÏANI, J. TIRADO. *Geology: Modular Georecommendation in Gossip-Based Social Networks*, in "International Conference on Distributed Computing Systems", Macau, China, 2012, http://hal.inria.fr/hal-00764234.

[26] T. CRAIN, V. GRAMOLI, M. RAYNAL. *A speculation-friendly binary search tree*, in "17th ACM SIGPLAN Symposium on Principles and Practice of Parallel Programming, PPOPP 2012", New Orleans, United States, 2012, http://hal.inria.fr/hal-00764358.

[27] T. CRAIN, V. GRAMOLI, M. RAYNAL. *Brief Announcement: A Contention-Friendly, Non-blocking Skip List*, in "Distributed Computing - 26th International Symposium, DISC 2012", Salvador, Brazil, 2012, http://hal.inria.fr/hal-00764360.

[28] T. CRAIN, D. IMBS, M. RAYNAL. *Towards a universal construction for transaction-based multiprocess programs*, in "13th International Conference on Distributed Computing and Networking (ICDCN'12)", Hong Kong, Hong Kong, Springer-Verlag LNCS, 2012, http://hal.inria.fr/hal-00646911.

[29] T. CRAIN, E. KANELLOU, M. RAYNAL. *STM Systems: Enforcing Strong Isolation between Transactions and Non-transactional Code*, in "Algorithms and Architectures for Parallel Processing - 12th International Conference, ICA3PP 2012", Fukuoka, Japan, 2012, http://hal.inria.fr/hal-00764356.

[30] D. FREY, A.-M. KERMARREC, K. KLOUDAS. *Probabilistic Deduplication for Cluster-Based Storage Systems*, in "ACM Symposium on Cloud Computing", San Jose, CA, United States, 2012, http://hal.inria.fr/hal-00728215.

[31] S. GAMBS, R. GUERRAOUI, H. HAMZA, F. HUC, A.-M. KERMARREC. *Scalable and Secure Polling in Dynamic Distributed Networks*, in "31st International Symposium on Reliable Distributed Systems (SRDS)", Irvine, California, United States, October 2012, http://hal.inria.fr/hal-00723566.

[32] 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.

[33] 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.

[34] K. HUGUENIN, A.-M. KERMARREC, K. KLOUDAS, F. TAÏANI. *Content and Geographical Locality in User-Generated Content Sharing Systems*, in "22nd SIGMM International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)", Toronto, Canada, June 2012, http://hal.inria.fr/hal-00686251.

[35] S. JIEKAK, A.-M. KERMARREC, N. LE SCOUARNEC, G. STRAUB, A. VAN KEMPEN. *Regenerating Codes: A System Perspective*, in "Dependability Issues in Cloud Computing (DISCCO 2012)", Irvine, California, United States, October 2012, p. 436-441 [*DOI :* 10.1109/SRDS.2012.58], http://hal.inria.fr/hal-00764262.

[36] A.-M. KERMARREC, E. LE MERRER. *Offline social networks: stepping away from the internet*, in "Proceedings of the Fifth Workshop on Social Network Systems", New York, NY, USA, SNS '12, ACM, 2012, p. 14:1–14:2, http://doi.acm.org/10.1145/2181176.2181190.

[37] A.-M. KERMARREC, E. LE MERRER, G. STRAUB, A. VAN KEMPEN. *Availability-based methods for distributed storage systems*, in "SRDS 2012, 31st International Symposium on Reliable Distributed Systems", Irvine, California., United States, October 2012, p. 151-160 [*DOI :* 10.1109/SRDS.2012.10], http://hal.inria.fr/hal-00521034.

[38] 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.

[39] A. MOSTEFAOUI, M. RAYNAL, J. STAINER. *Chasing the Weakest Failure Detector for k-Set Agreement in Message-passing Systems*, in "NCA 2012 - 11th Annual IEEE International Symposium On Network Computing and Applications", Boston, United States, 2012, p. 44-51, http://hal.inria.fr/hal-00733088.

[40] M. RAYNAL, J. STAINER. *A Simple Asynchronous Shared Memory Consensus Algorithm Based on Omega and Closing Sets*, in "CISIS 2012 - Sixth International Conference on Complex, Intelligent, and Software Intensive Systems", Palerme, Italy, 2012, p. 357-364, http://hal.inria.fr/hal-00733082.

[41] M. RAYNAL, J. STAINER. *From a Store-Collect Object and Ω to Efficient Asynchronous Consensus*, in "Euro-Par - Parallel Processing - 18th International Conference - 2012", Rhodes Island, Greece, 2012, p. 427-438, http://hal.inria.fr/hal-00733080.

[42] M. RAYNAL, J. STAINER. *Increasing the Power of the Iterated Immediate Snapshot Model with Failure Detectors*, in "SIROCCO - Structural Information and Communication Complexity - 19th International Colloquium - 2012", Reykjavik, Iceland, 2012, p. 231-242, http://hal.inria.fr/hal-00733077.

National Conferences with Proceeding

[43] F. LE FESSANT, T. GAZAGNAIRE. *Ocp-build: un gestionnaire de projets pour OCaml*, in "JFLA - Journées Francophones des Langages Applicatifs - 2012", Carnac, France, February 2012, http://hal.inria.fr/hal-00665962.

Research Reports

[44] T. CRAIN, V. GRAMOLI, M. RAYNAL. *A Contention-Friendly Methodology for Search Structures*, University of Rennes 1, February 2012, http://hal.inria.fr/hal-00668010.

[45] T. CRAIN, V. GRAMOLI, M. RAYNAL. *A Contention-Friendly, Non-Blocking Skip List*, Inria, May 2012, n^{o} RR-7969, http://hal.inria.fr/hal-00699794.

[46] T. CRAIN, E. KANELLOU, M. RAYNAL. *STM systems: Enforcing strong isolation between transactions and non-transactional code*, Inria, May 2012, n^{o} RR-7970, http://hal.inria.fr/hal-00699903.

[47] A.-M. KERMARREC, A. MOIN. *Data Visualization Via Collaborative Filtering*, Inria, February 2012, 23, http://hal.inria.fr/hal-00673330.

[48] A.-M. KERMARREC, A. MOIN. *FlexGD : A Flexible Force-directed Model for Graph Drawing*, Inria, November 2012, 13, http://hal.inria.fr/hal-00679574.

[49] M. RAYNAL, J. STAINER. *Increasing the Power of the Iterated Immediate Snapshot Model with Failure Detectors*, University of Rennes 1, February 2012, n^{o} PI-1991, http://hal.inria.fr/hal-00670154.

References in notes

[50] 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.

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

[52] 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, p. 41-88.

[53] 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.

[54] C. DWORK. *Differential privacy: a survey of results*, in "TAMC", 2008.

[55] 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.

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

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

[58] 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.

[59] 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.

[60] 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.

[61] 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.