IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Rennes

Université Rennes 1

Activity Report 2014

Project-Team ASAP

As Scalable As Possible: foundations of large scale dynamic distributed systems

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Distributed Systems and middleware

Table of contents

Project-Team ASAP

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

Creation of the Project-Team: 2007 July 01.

1. Members

Research Scientists

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

Faculty Members

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

Engineers

Yuri Barssi [Univ. Rennes I, until Jun 2014, granted by FP7 EIT ICT LABS GA project]
Heverson Borba Ribeiro [Inria, until Jun 2014, granted by FP7 EIT ICT LABS GA project]
Antoine Boutet [Inria, until Jan 2014, granted by FP7 EIT ICT LABS GA project]
Raziel Carvajal Gomez [Inria, granted by Brow2Brow ADT]
Jacques Falcou [Inria, until May 2014, granted by FP7 EIT ICT LABS GA project]
Arnaud Jégou [Inria, granted by FP7 EIT ICT LABS GA project]
Xavier Lucas [Inria, from Oct 2014]
Jean-François Verdonck [Inria, granted by FP7 EIT ICT LABS GA project]

PhD Students

Fabien André [Technicolor, granted by CIFRE]
Resmi Ariyattu Chandrasekharannair [Univ. Rennes I]
Simon Bouget [Normale Sup Rennes, from Sep 2014]
Stéphane Delbruel [Univ. Rennes I]
Ali Gouta [Orange Labs, granted by CIFRE]
Nupur Mittal [Inria, granted by Google Focused Award Alter-Egos]
Javier Olivares [Univ. Federico Santa Maria Valparaiso, external grant]
Antoine Rault [Inria, granted by Conseil Régional de Bretagne]
Pierre-Louis Roman [Univ. Rennes I, from Oct 2014]
Julien Stainer [Univ. Rennes I]

Post-Doctoral Fellows

Zohir Bouzid [Univ. Rennes I, from Jun 2014]
Nitin Chiluka [Univ. Rennes I]
Juan Manuel Tirado Martin [Inria]

Visiting Scientists

Yahya Benkaouz [Inria, until Feb 2014]
Maryam Helmi Khomeirani [Inria, from Aug 2014 until Sep 2014]
Frederik Mallmann-Trenn [Inria, until Jul 2014]
Diogo Saraiva Lima [Univ. Rennes I, from Jul 2014 until Aug 2014]

Administrative Assistant

Cécile Bouton [Inria, AER]

Others

Naman Goel [Inria, Internship, from Feb 2014 until Apr 2014, granted by Google Focused Award Alter-Egos]
Mathieu Pasquet [Univ. Rennes I, Internship, from Feb 2014 until Jun 2014]
Martin Sansoucy [Univ. Rennes I, Internship, from May 2014 until Aug 2014]

2. Overall Objectives

2.1. General objectives

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

2.1.1. Scalability

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

2.1.2. Personalization

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

2.1.3. Uncertainty

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

2.1.4. Malicious behaviors and privacy

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

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

2.2. Structure of the team

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

2.2.1. Objective 1: Decentralized personalization

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

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

2.2.2. Objective 2: Personalization: Cloud computing meets p2p

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

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

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

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

2.2.3. Objective 3: Privacy-aware decentralized computations

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

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

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

2.2.4. Objective 4: Information dissemination over social networks

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

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

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

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

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

2.2.5. Objective 5: Computability and efficiency of distributed systems

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

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

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

3. Research Program

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*" [55] (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 [56], [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 thousands or even millions of computing entities scattered all over the world and dealing with a huge amount of data. This major shift in scalability requirements has lead to the emergence of novel computing paradigms. In particular, the peer-to-peer communication paradigm imposed itself as the prevalent model to cope with the requirements of large scale distributed systems. Peer-to-peer systems rely on a symmetric communication model where peers are potentially both clients and servers. They are fully decentralized, thus avoiding the bottleneck imposed by the presence of servers in traditional systems. They are highly resilient to peers arrivals and departures. Finally, individual peer behavior is based on a local knowledge of the system and yet the system converges toward global properties.

A peer-to-peer overlay network logically connects peers on top of IP. Two main classes of such overlays dominate, structured and unstructured. The differences relate to the choice of the neighbors in the overlay, and the presence of an underlying naming structure. Overlay networks represent the main approach to build large-scale distributed systems that we retained. An overlay network forms a logical structure connecting participating entities on top of the physical network, be it IP or a wireless network. Such an overlay might form a structured overlay network [57], [58], [59] following a specific topology or an unstructured network [54], [60] 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 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 an arbitrary value. In this case, the validity property of consensus, which is normally "a decided value is a proposed value", must be changed to "if all correct processes propose the same value then only this value can be decided." Moreover, the resilience bound of less than half of faulty processes is at least lowered to "less than a third of Byzantine processes." These are some of the aspects that underlie our studies in the context of the classical model of distributed systems, in peer-to-peer systems and in sensor networks.

3.6. Online social networks and recommender systems

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.

We also investigate approaches to build implicit social networks, connecting users sharing similar interests. At the heart of the building of such similarity graphs lie k-nearest neighbor (KNN) algorithms. Our research in this area is to design and implement efficient KNN algorithms able to cope with a huge volume of data as well as a high level of dynamism. We investigate the use of such similarity graphs to build highly scalable infrastructures for recommendation systems.

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. New Software and Platforms

5.1. MediEgo: A recommendation solution for webmasters

Participants: Jacques Falcou, Arnaud Jégou, Xavier Lucas, Anne-Marie Kermarrec, Jean-François Verdonck.

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

MEDIEGO is a solution for content recommendation based on the users navigation history. The solution 1) collects the usages of the Web users and store them in a profile; 2) uses this profile to associate to each user her most similar users; 3) leverages this implicit network of close users in order to infer their preferences and recommend advertisements and recommendations. MEDIEGO achieves scalability using a sampling method, which provides very good results at a drastically reduced cost. The MEDIEGO recommendation engine is built in collaboration with Sébastien Campion. We have demonstrated the software at the conference Le Web 14 (9-11 Dec 2014) in collaboration with France Television.

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

Participants: Heverson Borba Ribeiro, Raziel Carvajal Gomez, Davide Frey, Arnaud Jégou, Anne-Marie Kermarrec.

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

Within the context of the AllYours EIT/ICT-Labs project, we refined the implementation of WhatsUp into the Peer-to-Peer AllYours application. AllYours-P2P is a peer-to-peer based news recommender system that organizes users into an implicit social network based on their explicit opinions. In AllYours-P2P the recommendation process is collaboratively performed by connected users. Every user runs a symmetric piece of software responsible for storing user interests and calculating the affinity between a user and its neighborhood. The local computed similarity then is used to keep virtual connections to other users whose interests are alike and remove connection to the ones that are not. As a result, an interest-based overlay is built and users converge to groups of similar interests within which news are disseminated. The AllYours-P2P software consists of two parts, running on each peer: an embedded application server, based on Jetty, and a web interface accessible from any web browser. The back-end is written in Java, while the user interface comprises HTML and Javascript code. AllYours-P2P is currently available in three different platforms: Mac OSx (10.5 or later), Windows (Vista and Windows 7) and Linux (Ubuntu 10.4 or later). We have tested Allyours-p2p in a real life environment with a set of invited users in Italy from Sep to Nov 2014. These test were a part of joint project between ASAP Team and its Italian partner Trentorise.

Currently the implementation of Allyours-p2p includes approximately 21K lines of code.

5.3. HyRec: A hybrid recommender system

Participants: Davide Frey, Anne-Marie Kermarrec.

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

This work leads to the development of HyRec, a hybrid recommender system. The motivation of this work is to explore solutions that could in some sense democratize personalization by making it accessible to any content provider company without generating huge investments. HyRec implements a user-based collaborative filtering scheme and offloads CPU-intensive recommendation tasks to front-end client browsers, while retaining storage and orchestration tasks within back-end servers. HyRec seeks to provide the scalability of p2p approaches without forcing content providers to give up the control of the system. This software has been developed in collaboration with Antoine Boutet (Univ. Saint-Étienne) and Rhicheek Patra (EPFL).

5.4. GossipLib: A library for gossip-based applications

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

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

GossipLib is a library consisting of a set of JAVA classes aimed to facilitate the development of gossip-based application in a large-scale setting. It provides developers with a set of support classes that constitute a solid starting point for building any gossip-based application. GossipLib is designed to facilitate code reuse and testing of distributed application, and provides also the implementation of a number of standard gossip protocols that may be used out of the box or extended to build more complex protocols and applications. These include for example the peer-sampling protocols for overlay management. GossipLib also provides facility for the configuration and deployment of applications as final-product but also as research prototype in environments like PlanetLab, clusters, network emulators, and even as event-based simulation. The code developed with GossipLib can be run both as a real application and in simulation. Currently the implementation of GossipLib includes approximately 9K lines of code, and is used in several projects by ASAP, including HEAP, AllYours-P2P, and Behave.

5.5. YALPS: A library for p2p applications

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

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

YALPS is an open-source Java library designed to facilitate the development, deployment, and testing of distributed applications. Applications written using YALPS can be run both in simulation and in real-world mode without changing a line of code or even recompiling the sources. A simple change in a configuration file will load the application in the proper environment. A number of features make YALPS useful both for the design and evaluation of research prototypes and for the development of applications to be released to the public. Specifically, YALPS makes it possible to run the same application as a simulation or in a real deployment. Applications communicate by means of application-defined messages which are then routed either through UDP/TCP or through YALPS's simulation infrastructure. In both cases, YALPS's communication layer offers features for testing and evaluating distributed protocols and applications. Communication channels can be tuned to incorporate message losses or to constrain their outgoing bandwidth. Finally, YALPS includes facilities to support operation in the presence of NATs and firewalls using relaying and NAT-traversal techniques. The implementation of YALPS includes approximately 16K lines of code, and is used in several projects by ASAP, including HEAP, AllYours-P2P, and Behave. This work was done in collaboration with Maxime Monod (EPFL).

5.6. HEAP: Heterogeneity-aware gossip protocol

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

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

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

5.7. Brow2Brow: Browser-to-browser serverless toolboxes

Participants: Raziel Carvajal Gomez, Davide Frey, Anne-Marie Kermarrec.

Brow2Brow is an "Action de Development Technologique", i.e. a collaborative development project that aims at providing a middleware and software library for browser-to-browser applications. Brow2Brow involves the ASAP team as well as the DICE Team from Inria Grenoble (Antenne de Lyon). The project seeks to provide an alternative to the current model followed by Web2.0 applications by exploiting the recently introduced WebRTC standard. Existing Web 2.0 applications collect data on browsers and send it to servers that store and process it. The goal of Brow2Brow is to provide an alternative approach where browsers can themselves proceed to collaborative data processing. This will make it possible avoid data concentration at a single server. The project has resulted so far in the development of WebGC, a library for gossip-based applications on browsers.

5.8. WebGC: Web-based Gossip Communication

Participants: Raziel Carvajal Gomez, Davide Frey, Anne-Marie Kermarrec.

Contact:	Raziel Carvajal Gomez, Davide Frey
License:	Not-yet released
Presentation:	Library for Gossip protocols within Web Browsers
Status:	Ongoing development

WebGC is a library for gossip-based communication between web-browsers. It has been developed in collaboration with Mathieu Simonin in the context of the Brow2Brow ADT project. WebGC builds on the recent WebRTC standard as well as on PeerJS, an open-source project that provides primitives for data transfer on top of WebRTC.

The library currently includes the implementation of two peer sampling protocols, CYCLON and the generic peer-sampling protocol from [7], as well as a clustering protocol [1]. All protocols implement a common GOSSIPPROTOCOL "interface"—since Javascript does not natively support interfaces, we adopt the interface pattern. A COORDINATOR makes it possible to stack these protocols on top of each other to implement applications.

6. New Results

6.1. Highlights of the Year

- Anne-Marie Kermarrec is the recipient of the **ACM/IFIP/USENIX/Middleware 10-Years Best Paper Award**, for her paper *The peer sampling service: Experimental evaluation of unstructured gossip-based implementations* (Middleware 2004), co-authored with Márk Jelasity, Rachid Guerraoui, and Maarten van Steen.
- Anne-Marie Kermarrec is the recipient of the **WISE 2014 Best Paper Award**, for her paper [18], co-authored with Alexandra Olteanu and Karl Aberer.
- Michel Raynal is the recipient of the **PODC 2014 Best Paper Award**, for his paper [34], co-authored with Achour Mostefaoui and Moumen Hamouna.
- The MEDIEGO recommendation engine was demonstrated at **Le Web 14** in partnership with FranceTV.

BEST PAPERS AWARDS :

[18] **Comparing the Predictive Capability of Social and Interest Affinity for Recommendations in 15th International Conference on Web Information System Engineering (WISE 2014)**. O. ALEXANDRA, A.-M. KERMARREC, K. ABERER.

[34] **Signature-Free Asynchronous Byzantine Consensus with t<n/3 and O(n2) Messages in ACM PODC**. A. MOSTEFAOUI, M. HAMOUNA, M. RAYNAL.

6.2. Models and abstractions for distributed systems

6.2.1. Signature-free asynchronous Byzantine consensus

Participant: Michel Raynal.

In [34] we present a new round-based asynchronous consensus algorithm that copes with up to $t < n/3$ Byzantine processes, where n is the total number of processes. In addition of not using signature, not assuming a computationally-limited adversary, while being optimal with respect to the value of t, this algorithm has several noteworthy properties: the expected number of rounds to decide is four, each round is composed of two or three communication steps and involves $O(n^2)$ messages, and a message is composed of a round number plus a single bit. To attain this goal, the consensus algorithm relies on a common coin as defined by Rabin, and a new extremely simple and powerful broadcast abstraction suited to binary values. The main target when designing this algorithm was to obtain a cheap and simple algorithm. This was motivated by the fact that, among the first-class properties, simplicity –albeit sometimes under-estimated or even ignored– is a major one.

This is a joint work with Achour Mostéfaouin and Hamouma Moumen. It received the PODC 2014 Best Paper Award.

6.2.2. Randomized mutual exclusion with constant amortized RMR complexity on the DSM

Participant: George Giakkoupis.

In [30] we settle an open question by determining the remote memory reference (RMR) complexity of randomized mutual exclusion, on the distributed shared memory model (DSM) with atomic registers, in a weak but natural (and stronger than oblivious) adversary model. In particular, we present a mutual exclusion algorithm that has constant expected amortized RMR complexity and is deterministically deadlock free. Prior to this work, no randomized algorithm with $o(\log n/\log\log n)$ RMR complexity was known for the DSM model. Our algorithm is fairly simple, and compares favorably with one by Bender and Gilbert (FOCS 2011) for the CC model, which has expected amortized RMR complexity $O(\log^2 \log n)$ and provides only probabilistic deadlock freedom.

This is a joint work with Philipp Woelfel (Univ. of Calgary, Canada).

6.2.3. Reliable shared memory abstraction on top of asynchronous Byzantine message-passing systems

Participants: Michel Raynal, Julien Stainer.

This work is on the construction and the use of a shared memory abstraction on top of an asynchronous message-passing system in which up to t processes may commit Byzantine failures. This abstraction consists of arrays of n single-writer/multi-reader atomic registers, where n is the number of processes. Differently from usual atomic registers which record a single value, each of these atomic registers records the whole history of values written to it. A distributed algorithm building such a shared memory abstraction it first presented. This algorithm assumes $t < n/3$, which is shown to be a necessary and sufficient condition for such a construction. Hence, the algorithm is resilient-optimal. Then we present distributed algorithms built on top of this shared memory abstraction, which cope with up to t Byzantine processes. The simplicity of these algorithms constitutes a strong motivation for such a shared memory abstraction in the presence of Byzantine processes. For a lot of problems, algorithms are more difficult to design and prove correct in a message-passing system than in a shared memory system. Using a protocol stacking methodology, the aim of the proposed abstraction is to allow an easier design (and proof) of distributed algorithms, when the underlying system is an asynchronous message-passing system prone to Byzantine failures.

This work was done in collaboration with Damien Imbs and Sergio Rajsbaum. It has been published in SIRROCCO [32] and as a technical report [43].

6.2.4. Distributed Universality

Participants: Michel Raynal, Julien Stainer.

A notion of a universal construction suited to distributed computing has been introduced by M. Herlihy in his celebrated paper “Wait-free synchronization” (ACM TOPLAS, 1991). A universal construction is an algorithm that can be used to wait-free implement any object defined by a sequential specification. Herlihy’s paper shows that the basic system model, which supports only atomic read/write registers, has to be enriched with consensus objects to allow the design of universal constructions. The generalized notion of a k-universal construction has been recently introduced by Gafni and Guerraoui (CONCUR 2011). A k-universal construction is an algorithm that can be used to simultaneously implement k objects (instead of just one object), with the guarantee that at least one of the k constructed objects progresses forever. While Herlihy’s universal construction relies on atomic registers and consensus objects, a k-universal construction relies on atomic registers and k-simultaneous consensus objects (which are wait-free equivalent to k-set agreement objects in the read/write system model). This work significantly extends the universality results introduced by Herlihy and Gafni-Guerraoui. In particular, we present a k-universal construction which satisfies the following five desired properties, which are not satisfied by the previous k-universal construction: (1) among the k objects that are constructed, at least ℓ objects (and not just one) are guaranteed to progress forever; (2) the progress condition for processes is wait-freedom, which means that each correct process executes an infinite number of operations on each object that progresses forever; (3) if any of the k constructed objects stops progressing, all its copies (one at each process) stop in the same state; (4) the proposed construction is contention-aware, in the sense that it uses only read/write registers in the absence of contention; and (5) it is generous with respect to the obstruction-freedom progress condition, which means that each process is able to complete any one of its pending operations on the k objects if all the other processes hold still long enough. The proposed construction, which is based on new design principles, is called a (k,ℓ)-universal construction. It uses a natural extension of k-simultaneous consensus objects, called (k,ℓ)-simultaneous consensus objects ((k,ℓ)-SC). Together with atomic registers, (k,ℓ)-SC objects are shown to be necessary and sufficient for building a (k,ℓ)-universal construction, and, in that sense, (k,ℓ)-SC objects are (k,ℓ)-universal.

This work was done in collaboration with Gadi Taubenfeld. It has been published as a brief announcement in PODC [37] and the full version appeared in OPODIS [38]. A version has also been published as a technical report [45].

6.2.5. *Computing in the presence of concurrent solo executions*

Participants: Michel Raynal, Julien Stainer.

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

This work was done in collaboration with Maurice Herlihy and Sergio Rajsbaum. It has been published in LATIN [31].

6.2.6. *A simple broadcast algorithm for recurrent dynamic systems*

Participants: Michel Raynal, Julien Stainer.

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

This work was done in collaboration with Jiannong Cao and Weigang Wu. It has been published in AINA [36].

6.2.7. *Fisheye consistency: Keeping data in synch in a georeplicated world*

Participants: Michel Raynal, François Taïani.

Over the last thirty years, numerous consistency conditions for replicated data have been proposed and implemented. Popular examples of such conditions include linearizability (or atomicity), sequential consistency, causal consistency, and eventual consistency. These consistency conditions are usually defined independently from the computing entities (nodes) that manipulate the replicated data; i.e., they do not take into account how computing entities might be linked to one another, or geographically distributed. To address this lack, as a first contribution, this work [41] introduces the notion of proximity graph between computing nodes. If two nodes are connected in this graph, their operations must satisfy a strong consistency condition, while the operations invoked by other nodes are allowed to satisfy a weaker condition. The second contribution is the use of such a graph to provide a generic approach to the hybridization of data consistency conditions into the same system. We illustrate this approach on sequential consistency and causal consistency, and present a model in which all data operations are causally consistent, while operations by neighboring processes in the proximity graph are sequentially consistent. The third contribution of this work is the design and the proof of a distributed algorithm based on this proximity graph, which combines sequential consistency and causal consistency (the resulting condition is called fisheye consistency). In doing so this work not only extends the domain of consistency conditions, but provides a generic provably correct solution of direct relevance to modern georeplicated systems.

This work was done in collaboration with Roy Friedman (The Technion, Haifa, Israel)

6.3. Large-scale and user-centric distributed systems

6.3.1. Archiving cold data in warehouses with clustered network coding

Participants: Fabien André, Anne-Marie Kermarrec.

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

In this work, we proposed a novel erasure code system, tailored for networked archival systems. The efficiency of our approach relies on the joint use of random codes and a clustered placement strategy. Our repair protocol leverages network coding techniques to reduce by 50% the amount of data transferred during maintenance, by repairing several cluster files simultaneously. We demonstrated both through an analysis and extensive experimental study conducted on a public testbed that our approach significantly decreases both the bandwidth overhead during the maintenance process and the time to repair lost data. We also showed that using a non-systematic code does not impact the throughput, and comes only at the price of a higher CPU usage. Based on these results, we evaluated the impact of this higher CPU consumption on different configurations of data coldness by determining whether the cluster's network bandwidth dedicated to repair or CPU dedicated to decoding saturates first.

This work has been conducted in collaboration with Erwan Le Merrer, Nicolas Le Scouarnec, Gilles Straub (Technicolor) and A. van Kempen (Univ. Nantes) and published in ACM Eurosys 2014 [19].

6.3.2. WebGC: Browser-based gossiping

Participants: Raziel Carvajal Gomez, Davide Frey, Anne-Marie Kermarrec.

The advent of browser-to-browser communication technologies like WebRTC has renewed interest in the peer-to-peer communication model. However, the available WebRTC code base still lacks important components at the basis of several peer-to-peer solutions. Through a collaboration with Mathieu Simonin from the Inria SED in the context of the Brow2Brow ADT project, we started to tackle this problem by proposing WebGC, a library for gossip-based communication between web browsers. Due to their inherent scalability, gossip-based, or epidemic protocols constitute a key component of a large number of decentralized applications. WebGC thus represents an important step towards their wider spread. We demonstrated a preliminary version of the library at Middleware 2014 [47].

6.3.3. Large-scale graph processing in datacenters with bandwidth guarantees

Participants: Nitin Chiluka, Anne-Marie Kermarrec.

Recent research has shown that the performance of data-intensive applications in multi-tenant datacenters can be severely impacted by each other's network usage. Starvation for network bandwidth in such datacenters typically results in significantly longer completion times for large-scale distributed applications. To address this concern, researchers propose bandwidth guarantees for all the virtual machines (VMs) initiated by each tenant in the datacenter in order to provide a predictable performance for their applications. In our work, we focus on large-scale graph processing in such datacenters. More specifically, given k VMs with their respective bandwidth constraints and a large graph, we perform a k-way partition on the graph such that the subsequent computation of various algorithms (e.g., PageRank, graph factorization) take minimal time.

6.3.4. Scaling KNN computation over large graphs on a PC

Participants: Nitin Chiluka, Anne-Marie Kermarrec, Javier Olivares.

Frameworks such as GraphChi and X-Stream are increasingly gaining attention for their ability to perform scalable computation on large graphs by leveraging disk and memory on a single commodity PC. These frameworks rely on the graph structure to remain the same for the entire pe- riod of computation of various algorithms such as PageRank and triangle counting. As a consequence, these frameworks are not applicable to algorithms that require the graph struc- ture to change during their computation. In this work, we focus on one such algorithm – K-Nearest Neighbors (KNN) – which is widely used in recommender systems. Our approach aims to minimize random accesses to disk as well as the amount of data loaded/unloaded from/to disk so as to better utilize the computational power, thus improving the algorithmic efficiency. The preliminary design and results of our approach appeared in Middleware 2014 [23].

6.3.5. Privacy-preserving distributed collaborative filtering

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

In collaboration with Antoine Boutet from the Univ. St Etienne, and Rachid Guerraoui from EPFL, we proposed a new mechanism to preserve privacy while leveraging user profiles in distributed recommender systems. Our mechanism relies on two contributions: (*i*) an original obfuscation scheme, and (*ii*) a randomized dissemination protocol. We showed that our obfuscation scheme hides the exact profiles of users without significantly decreasing their utility for recommendation. In addition, we precisely characterized the conditions that make our randomized dissemination protocol differentially private.

We compared our mechanism with a non-private as well as with a fully private alternative. We considered a real dataset from a user survey and report on simulations as well as planetlab experiments. In short, our extensive evaluation showed that our twofold mechanism provides a good trade-off between privacy and accuracy, with little overhead and high resilience.

6.3.6. Behave: Behavioral cache for web content

Participants: Davide Frey, Anne-Marie Kermarrec.

In collaboration with Mathieu Goessens, a former intern of the team, we proposed Behave: a novel approach for peer-to-peer cache-oriented applications such as CDNs. Behave relies on the principle of Behavioral Locality inspired from collaborative filtering. Users that have visited similar websites in the past will have local caches that provide interesting content for one another.

Behave exploits epidemic protocols to build overlapping communities of peers with similar interests. Peers in the same one-hop community federate their cache indexes in a Behavioral cache. Extensive simulations on a real data trace show that Behave can provide zero-hop lookup latency for about 50% of the content available in a DHT-based CDN. The results of this work were published at DAIS 2014 [26].

6.3.7. HyRec: Leveraging browsers for scalable recommenders

Participants: Davide Frey, Anne-Marie Kermarrec.

The ever-growing amount of data available on the Internet calls for personalization. Yet, the most effective personalization schemes, such as those based on collaborative filtering (CF), are notoriously resource greedy. In this work, we proposed HyRec, an online cost-effective scalable system for user-based CF personalization. HyRec offloads recommendation tasks onto the web browsers of users, while a server orchestrates the process and manages the relationships between user profiles.

We fully implemented HyRec and we extensively evaluated it on several workloads from MovieLens and Digg. Our experiments conveyed the ability of HyRec to reduce the operation costs of content providers by nearly 50% and to provide a 100-fold improvement in scalability with respect to a centralized (or cloud-based recommender approach), while preserving the quality of personalization. HyRec is also virtually transparent to users and induces only 3% of the bandwidth consumption of a p2p solution. This work was done in collaboration with Antoine Boutet from the Univ. St Etienne, as well as with Rachid Guerraoui, and Rhicheek Patra from EPFL. It resulted in a publication at Middleware 2014 [22].

6.3.8. Landmark-based similarity for p2p collaborative filtering

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

Computing k-nearest-neighbor graphs constitutes a fundamental operation in a variety of data-mining applications. As a prominent example, user-based collaborative-filtering provides recommendations by identifying the items appreciated by the closest neighbors of a target user. As this kind of applications evolve, they will require KNN algorithms to operate on more and more sensitive data. This has prompted researchers to propose decentralized peer-to-peer KNN solutions that avoid concentrating all information in the hands of one central organization. Unfortunately, such decentralized solutions remain vulnerable to malicious peers that attempt to collect and exploit information on participating users.

We seek to overcome this limitation by proposing *H&S* (Hide & Share), a novel landmark-based similarity mechanism for decentralized KNN computation. Landmarks allow users (and the associated peers) to estimate how close they lay to one another without disclosing their individual profiles.

We evaluate *H&S* in the context of a user-based collaborative-filtering recommender with publicly available traces from existing recommendation systems. We show that although landmark-based similarity does disturb similarity values (to ensure privacy), the quality of the recommendations is not as significantly hampered. We also show that the mere fact of disturbing similarity values turns out to be an asset because it prevents a malicious user from performing a profile reconstruction attack against other users, thus reinforcing users' privacy. Finally, we provide a formal privacy guarantee by computing the expected amount of information revealed by *H&S* about a user's profile.

This work was done in collaboration with Jingjing Wang, and Rachid Guerraoui.

6.3.9. Adaptation for the masses: Towards decentralized adaptation in large-scale p2p recommenders

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

Decentralized recommenders have been proposed to deliver privacy-preserving, personalized and highly scalable on-line recommendation services. Current implementations tend, however, to rely on hard-wired, mechanisms that cannot adapt. Deciding beforehand which hard-wired mechanism to use can be difficult, as the optimal choice might depend on conditions that are unknown at design time. In [27], we have proposed a framework to develop dynamically adaptive decentralized recommendation systems. Our proposal supports a decentralized form of adaptation, in which individual nodes can independently select, and update their own recommendation algorithm, while still collectively contributing to the overall system's services.

This work was done in collaboration with Christopher Maddock and Andreas Mauthe (Univ. of Lancaster, UK).

6.3.10. Tight bounds for rumor spreading with vertex expansion

Participant: George Giakkoupis.

In [28] we establish an upper bound for the classic PUSH-PULL rumor spreading protocol on general graphs, in terms of the vertex expansion of the graph. We show that $O(\log^2(n)/\alpha)$ rounds suffice with high probability to spread a rumor from any single node to all n nodes, in any graph with vertex expansion at least α. This bound matches a known lower bound, and settles the natural question on the relationship between rumor spreading and vertex expansion asked by Chierichetti, Lattanzi, and Panconesi (SODA 2010). Further, some of the arguments used in the proof may be of independent interest, as they give new insights, for example, on how to choose a small set of nodes in which to plant the rumor initially, to guarantee fast rumor spreading.

6.3.11. Greedy routing in small-world networks with power-law degrees

Participant: George Giakkoupis.

In [12] we study decentralized routing in small-world networks that combine a wide variation in node degrees with a notion of spatial embedding. Specifically, we consider a variant of J. Kleinberg's grid-based small-world model in which (1) the number of long-range edges of each node is not fixed, but is drawn from a power-law probability distribution with exponent parameter $\alpha \geq 0$ and constant mean, and (2) the long-range edges are considered to be bidirectional for the purposes of routing. This model is motivated by empirical observations indicating that several real networks have degrees that follow a power-law distribution. The measured power-law exponent α for these networks is often in the range between 2 and 3. For the small-world model we consider, we show that when $2 < \alpha < 3$ the standard greedy routing algorithm, in which a node forwards the message to its neighbor that is closest to the target in the grid, finishes in an expected number of $O(\log^{\alpha-1} n \cdot \log\log n)$ steps, for any source–target pair. This is asymptotically smaller than the $O(\log^2 n)$ steps needed in Kleinberg's original model with the same average degree, and approaches $O(\log n)$ as α approaches 2. Further, we show that when $0 \leq \alpha < 2$ or $\alpha \geq 3$ the expected number of steps is $O(\log^2 n)$, while for $\alpha = 2$ it is $O(\log^{4/3} n)$. We complement these results with lower bounds that match the upper bounds within at most a $\log\log n$ factor.

This is a joint work with Pierre Fraigniaud (Inria Paris-Rocquencourt and CNRS).

6.3.12. Randomized rumor spreading in dynamic graphs

Participant: George Giakkoupis.

In [29] we consider the well-studied rumor spreading model in which nodes contact a random neighbor in each round in order to push or pull the rumor. Unlike most previous works which focus on static topologies, we look at a dynamic graph model where an adversary is allowed to rewire the connections between vertices before each round, giving rise to a sequence of graphs, $G_1, G_2, \ldots$ Our first result is a bound on the rumor spreading time in terms of the conductance of those graphs. We show that if the degree of each node does not change much during the protocol (that is, by at most a constant factor), then the spread completes within t rounds for some t such that the sum of conductances of the graphs G_1 up to G_t is $O(\log n)$. This result holds even against an adaptive adversary whose decisions in a round may depend on the set of informed vertices before the round, and implies the known tight bound with conductance for static graphs. Next we show that for the alternative expansion measure of vertex expansion, the situation is different. An adaptive adversary can delay the spread of rumor significantly even if graphs are regular and have high expansion, unlike in the static graph case where high expansion is known to guarantee fast rumor spreading. However, if the adversary is oblivious, i.e., the graph sequence is decided before the protocol begins, then we show that a bound close to the one for the static case holds for any sequence of regular graphs.

This is a joint work with Thomas Sauerwald (Univ. of Cambridge, UK) and Alexandre Stauffer (Univ. of Bath, UK).

6.3.13. Privacy-preserving dissemination in social networks and microblogs

Participants: George Giakkoupis, Arnaud Jégou, Anne-Marie Kermarrec, Nupur Mittal.

Online micro-blogging services and social networks, as exemplified by Twitter and Facebook, have emerged as an important means of disseminating information quickly and at large scale. A standard mechanism in micro-blogging that allows for interesting content to reach a wider audience is that of *reposting* (i.e., *retweeting* in Twitter, or *sharing* in Facebook) of content initially posted by another user. Motivated by recent events in which users were prosecuted merely for reposting anti-government information, we present in [42] Riposte, a randomized reposting scheme that provides privacy guarantees against such charges. The idea is that if the user likes a post, Riposte will repost it only with some (carefully chosen) probability; and if the user does not like it, Riposte may still repost it with a slightly smaller probability. These probabilities are computed for each user as a function of the number of connections of the user in the network, and the extent to which the post has already reached those connections. The choice of these probabilities is based on results for branching processes, and ensures that interesting posts (liked by a large fraction of users) are likely to disseminate widely, whereas uninteresting posts (or spam) do not spread. Riposte is executed locally at the user, thus the user's opinion on the post is not communicated to the micro-blogging server. In this work, we quantify Riposte's ability to protect users in terms of differential privacy and provide analytical bounds on the dissemination of posts. We also do extensive experiments based on topologies of real networks, including Twitter, Facebook, Renren, Google+ and LiveJournal.

This work has been carried out in collaboration with Rachid Guerraoui (EPFL).

6.3.14. Adaptive streaming

Participants: Ali Gouta, Anne-Marie Kermarrec.

HTTP Adaptive Streaming (HAS) is gradually being adopted by Over The Top (OTT) content providers. In HAS, a wide range of video bitrates of the same video content are made available over the internet so that clients' players pick the video bitrate that best fit their bandwidth. Yet, this affects the performance of some major components of the video delivery chain, namely CDNs or transparent caches since several versions of the same content compete to be cached. In this context, we investigated the benefits of a Cache Friendly HAS system (CF-DASH), which aims to improve the caching efficiency in mobile networks and to sustain the quality of experience of mobile clients. We conducted our work by presenting a set of observations we made on a large number of clients requesting HAS contents. We introduced the CF-Dash system and our testbed implementation. Finally, we evaluated CF-dash based on trace-driven simulations and testbed experiments. Our validation results are promising. Simulations on real HAS traffic show that we achieve a significant gain in hit-ratio that ranges from 15% up to 50%. This work was done in collaboration with Zied Aouini, Yannick Le Louedec and Diallo Mamadou, and was published in NOSSDAV 2014 [39].

6.3.15. Predictive capabilities of social and interest affinity for recommendations

Participant: Anne-Marie Kermarrec.

The advent of online social networks created new prediction opportunities for recommender systems: instead of relying on past rating history through the use of collaborative filtering (CF), they can leverage the social relations among users as a predictor of user tastes similarity. Alas, little effort has been put into understanding when and why (e.g., for which users and what items) the social affinity (i.e., how well connected users are in the social network) is a better predictor of user preferences than the interest affinity among them as algorithmically determined by CF, and how to better evaluate recommendations depending on, for instance, what type of users a recommendation application targets. This overlook is explained in part by the lack of a systematic collection of datasets including both the explicit social network among users and the collaborative annotated items. In this work, we conducted an extensive empirical analysis on six real-world publicly available datasets, which dissects the impact of user and item attributes, such as the density of social ties or item rating patterns, on the performance of recommendation strategies relying on either the social ties or past rating similarity. Our findings represent practical guidelines that can assist in future deployments and mixing schemes. This work has been done in collaboration with Karl Aberer and Alexandra Olteanu (EPFL Swizerland). The paper received the Best Paper Award at the WISE International Conference [18].

6.3.16. Polystyrene: The decentralized data shape that never dies

Participants: Anne-Marie Kermarrec, François Taïani.

Decentralized topology construction protocols organize nodes along a predefined topology (e.g. a torus, ring, or hypercube). Such topologies have been used in many contexts ranging from routing and storage systems, to publish-subscribe and event dissemination. Since most topologies assume no correlation between the physical location of nodes and their positions in the topology, they do not handle catastrophic failures well, in which a whole region of the topology disappears. When this occurs, the overall shape of the system typically gets lost. This is highly problematic in applications in which overlay nodes are used to map a virtual data space, be it for routing, indexing or storage. In this work [20], we propose a novel decentralized approach that maintains the initial shape of the topology even if a large (consecutive) portion of the topology fails. Our approach relies on the dynamic decoupling between physical nodes and virtual ones enabling a fast reshaping. For instance, our results show that a 51,200-node torus converges back to a full torus in only 10 rounds after 50% of the nodes have crashed. Our protocol is both simple and flexible and provides a novel form of collective survivability that goes beyond the current state of the art.

This work has been done in collaboration with Simon Bouget (ENS Rennes) and Hoel Kervadec (INSA Rennes).

6.3.17. Link-prediction for very large scale graphs using distributed graph engines

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

In this project, we consider how the emblematic problem of link-prediction can be implemented efficiently in gather-apply-scatter (GAS) platforms, a popular distributed graph-computation model. Our proposal, called SNAPLE, exploits a novel highly-localized vertex scoring technique, and minimizes the cost of data flow while maintaining prediction quality. When used within GraphLab, SNAPLE can scale to extremely large graphs that a standard implementation of link prediction on cannot handle within the same platform. More precisely, we show that our approach can process a graph containing 1.4 billions edges on a 256 cores cluster in less than three minutes, with no penalty in the quality of predictions. This result corresponds to an over-linear speedup of 30 against a 20-core stand-alone machine running a non-distributed state-of-the-art solution.

6.3.18. GOSSIPKIT: A unified component framework for gossip

Participant: François Taïani.

Although the principles of gossip protocols are relatively easy to grasp, their variety can make their design and evaluation highly time consuming. This problem is compounded by the lack of a unified programming framework for gossip, which means developers cannot easily reuse, compose, or adapt existing solutions to fit their needs, and have limited opportunities to share knowledge and ideas. In [17], we have considered how component frameworks, which have been widely applied to implement middleware solutions, can facilitate the development of gossip-based systems in a way that is both generic and simple. We show how such an approach can maximise code reuse, simplify the implementation of gossip protocols, and facilitate dynamic evolution and re-deployment.

This work was done in collaboration with Shen Lin (SAP Labs) and Gordon Blair (Univ. of Lancaster, UK).

6.3.19. Towards a new model for cyber foraging

Participant: François Taïani.

Cyber foraging seeks to expand the capabilities and battery life of mobile devices by offloading intensive computations to nearby computing nodes (the surrogates). Although promising, current approaches to cyber foraging tend to impose a strict separation between the application state maintained on the mobile device, and data processed on the surrogates. In [33], we argue that this separation limits the applicability of cyber foraging, and explore how state sharing could be implemented in practice.

This work was done in collaboration with Diogo Lima and Hugo Miranda (Univ. of Lisbon, Portugal).

7. Bilateral Contracts and Grants with Industry

7.1. Technicolor

Participants: Fabien André, Anne-Marie Kermarrec.

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

7.2. Orange Labs

Participants: Ali Gouta, Anne-Marie Kermarrec.

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

7.3. Web Alter-Egos Google Focused Award

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

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

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

8. Partnerships and Cooperations

8.1. National initiatives

8.1.1. LABEX CominLabs

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

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

8.1.2. ANR project SocioPlug

Participants: Davide Frey, Anne-Marie Kermarrec, Pierre-Louis Roman, François Taïani.

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

8.1.3. DeSceNt CominLabs

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

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

8.1.4. ANR Blanc project Displexity

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

The Displexity project started in Oct 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 & H2020 projects

8.2.1.1. TOWARD THE ALLYOURS START-UP

Title: TOWARD THE ALLYOURS START-UP

Type: EIT-ICT Labs

Instrument: ACLD Computing in the Cloud

Duration: Jan - Dec 2014.

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. In this second year, the AllYours activity focused on the peer-to-peer and on the corporate version of AllYours through a collaborative initiative that involves the ASAP team, TrentoRise (Italy), and the Eindhoven EIT/ICT nodes. Our work consisted on refining and testing our implementations. For the p2p version, we ran a test with real users coordinated by TrentoRise from Sep to Nov 2014.

8.2.2. Collaborations with major European organizations

Ecole Polytechnique Federale de Lausanne EPFL Switzerland; collaboration on the Google Focused Award Web-Alter-Egos.

8.3. International Initiatives

8.3.1. Inria associate teams

8.3.1.1. RADCON

Title: Randomized Algorithms for Distributed Computing and Networks

International Partner (Institution - Laboratory - Researcher):

Univ. of Calgary (CANADA)

Duration: 2013 - 2015

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

Over recent years, computing systems have seen a massive increase in parallelism and interconnectivity. Peer-to-peer systems, ad-hoc networks, sensor networks, or the "cloud" are based on highly connected and volatile networks. Individual nodes such as cell phones, desktop computers or high performance computing systems rely on parallel processing power achieved through multiple processing units. To exploit the power of massive networks or multiple processors, algorithms must cope with the scale and asynchrony of these systems, and their inherent instability, e.g., due to node, link, or processor failures. In this research project we explore randomized algorithms for large-scale networks of distributed systems, and for shared memory multi-processor systems.

For large-scale networks, decentralized gossip protocols have emerged as a standard approach to achieving fault-tolerant communication between nodes with simple and scalable algorithms. We will devise new gossip protocols for various complex distributed tasks, and we will explore the power and limits of gossip protocols in various settings.

For shared memory systems, randomized algorithms have proved extremely useful to deal with asynchrony and failures. Sometimes probabilistic algorithms provide the only solution to a problem; sometimes they are more efficient; sometimes they are simply easier to implement. We will devise efficient algorithms for some of the fundamental problems of shared memory computing, such as mutual exclusion, renaming, and consensus.

8.3.2. Inria international partners

Univ. of Calgary

Univ. Nacional Autonoma de Mexico

Univ. of Glasgow

8.4. International Research Visitors

8.4.1. Visits of international scientists

Yahya Benkaouz, ENSIAS Rabat, Morocco, from Dec 1 2013 to Feb 28 2014

Maryam Helmi Khomeirani, Univ. of Calgary, Canada, from Aug 15 to Oct 14 2014

Frederik Mallmann-Trenn, Simon Fraser Univ., Canada, from Jan 15 to Apr 20 and from Jun 16 to Jul 22 2014

Diogo Saraiva Lima, Univ. of Lisbon, Portugal, from Jul 1 to Aug 29 2014

8.4.2. Internships

Naman Goel; from Feb 1 to Apr 25 2014. *User profiles segmentation for efficient personalized recommendations*. Supervised by Anne-Marie Kermarrec and François Taiani.

Frederik Mallmann-Trenn; until Jul 31 2014. *Bounds on the Voting Time in Terms of Conductance.* Supervised by George Giakkoupis (and Petra Berenbrink of Simon Frazer Univ., Canada).

Mathieu Pasquet; from Feb 1 to Jun 30 2014. *Content-based orientation in decentralized recommenders.* Supervised by Davide Frey.

Martin Sansoucy; from May 15 to Aug 31 2014. *Caractérisation d'un protocole décentralisé de construction de topologies informatiques réparties à mémoire de forme*. Supervised by François Taiani.

8.4.3. Visits to international teams

George Giakkoupis, Univ. of Calgary, Canada, Mar 23 to Apr 10, and Oct 22 to Nov 9 2014.

Anne-Marie Kermarrec was a part-time visiting professor at EPFL, Switzerland, from Jan to Jul 2014.

Anne-Marie Kermarrec, Univ. of Sydney and NICTA, Australia, two weeks in Jan 2014.

Anne-Marie Kermarrec, Yandex, Moscow, Russia, one week in Jun 2014.

Antoine Rault, EPFL, Switzerland, Sep 1 to Nov 29 Nov 2014

9. Dissemination

9.1. Promoting scientific activities

Davide Frey served in the program committees of the following conferences:

13th IEEE International Conference on Peer-to-Peer Computing (P2P), London, England, Sep 2014.

ACM/IFIP/USENIX International Conference on Middleware (Middleware), Bordeaux, France, Dec 2014.

Middleware for Next Generation Internet Computing (MW4NG), Bordeaux, France, Dec 2014.

34th International Conference on Distributed Computing Systems (ICDCS), Madrid, Spain, Jul 2014.

IEEE International Parallel & Distributed Processing Symposium (IPDPS), Phoenix, Arizona, USA, May 2014.

George Giakkoupis organized the *Workshop on Randomized Algorithms for Distributed Computing and Networks (RADICON)*, Rennes, France, Jul 2014.

George Giakkoupis served in the program committees of the following conferences:

21st International Colloquium on Structural Information and Communication Complexity (SIROCCO), Takayama, Japan, Jul 2014.

28th International Symposium on Distributed Computing (DISC), Austin, TX, USA, Oct 2014.

George Giakkoupis served in the conference committee of the *ACM Symposium on Principles of Distributed Computing (PODC)*, Paris, France, Jul 2014.

Anne-Marie Kermarrec is a member of the *Academia Europea* since 2013.

Anne-Marie Kermarrec is a member of the scientific committee of the *Société Informatique de France*.

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

Anne-Marie Kermarrec was the chair of the *ACM Software System Award Committee* in 2014.

Anne-Marie Kermarrec is a member of the *ACM Software System Award Committee* since 2009.

Anne-Marie Kermarrec is a member of the *ERC panel for Consolidator Grants* since 2013.

Anne-Marie Kermarrec is an associate editor of *IEEE Internet Computing*.

Anne-Marie Kermarrec is an associate editor of the *Springer Computing Journal*.

Anne-Marie Kermarrec was a member of the hiring committee at EPFL in 2014.

Anne-Marie Kermarrec is the vice-chair of *ACM Eurosys*.

Anne-Marie Kermarrec served in the program committees of the following conferences:

International Conference on Distributed Computing Systems (ICDCS), Madrid, Jul 2014.

International Conference on Extending Database Technology (EDBT), Athens, Greece, Mar 2014.

International Colloquium on Automata, Languages and Programming (ICALP), Copenhaguen, Sweden, Jul 2014.

ACM/IFIP/USENIX International Conference on Middleware (Middleware), Bordeaux, France, Dec 2014.

International Conference on Networked Systems (NETYS), Marrakech, Morocco, May 2014.

ACM Eurosys, Amsterdam, The Netherlands, Apr 2014.

ACM International Systems and Storage Conference (SYSTOR), Haifa, Israel, Jun 2014.

International Conference on Distributed Computing Systems (ICDCS), Madrid, Jul 2014.

International Conference on Extending Database Technology (EDBT), Athens, Greece, Mar 2014.

Anne-Marie Kermarrec is a member of the steering committees of *ACM Eurosys*, *Middleware*, and the *Winter School on Hot Topics in Distributed Systems*.

Anne-Marie Kermarrec is a member of the technical committee of the *ACM Conference on Online Social Networks*.

Anne-Marie Kermarrec co-organized the *Workshop Inria/Technicolor on Distributed Storage Systems*, Rennes, Dec 2014.

Anne-Marie Kermarrec co-organized the *Web Alter-Ego Workshop*, EPFL, Lausanne, Swizerland, Jan 2014.

Anne-Marie Kermarrec participated in the Forum Liberation de Rennes, Apr 2014.

Anne-Marie Kermarrec wrote two articles on the gender gap in computer science: (i) *Informatique [nom feminin]*, on the blog of the SIF Binaire on Mar 2014 (http://binaire.blog.lemonde.fr/2014/03/07/informatique-nom-feminin-2/); (ii) *Computer Science: Too young to fall into the gender gap*, IEEE Internet Computing [15].

Michel Raynal was in the editorial board of the journals: *IEEE Transactions on Computers*, *Journal of Parallel and Distributed Computing*, *Journal of Computer Systems Science and Engineering*, *Foundations of Computing and Decision Sciences*.

Michel Raynal served as a PC co-chair of the *International Conference on Networked Systems (NETYS)*, Marrakech, Morocco, May 2014.

Michel Raynal served in the program committee of the *ACM Symposium on Principles of Distributed Computing (PODC)*, Paris, France, Jul 2014.

Michel Raynal was the Award chair of the *IEEE Symposium on Reliable Distributed Systems (SRDS)* 2014.

Michel Raynal was a *European Representative in the IEEE Technical Committee on Distributed Computing*.

Michel Raynal was an international liaison co-chair of *ICDCS 2013*.

François Taïani served in the program committees of the following conferences:

34rd International Conference on Distributed Computing Systems (ICDCS), Madrid, Spain, Jun 2014.

33rd IEEE Symposium on Reliable Distributed Systems (SRDS), Nara, Japan, Oct 2014.

14th International IFIP Conference on Distributed Applications and Interoperable Systems (DAIS), Berlin, Germany, Jun 2014.

François Taïani served as PC Co-Chair of the *15th ACM/IFIP/USENIX Middleware Conference (Middleware)*, Dec 2014, Bordeaux, France.

François Taïani is a member of the Steering Committee of the *International IFIP Conference on Distributed Applications and Interoperable Systems (DAIS)* since 2013.

Juan Manuel Tirado Martin was a program co-chair of the conferences:

17th IEEE International Conference on Computation Science and Engineering (CSE), Chengdu, China, 2014.

6th IEEE International Conference on Cloud Computing Technology and Science (CloudCom), Singapore, 2014.

Middleware for Next Generation Internet Computing (MW4NG), Bourdeaux, France, 2014.

9.2. Invited talks

George Giakkoupis. *Rumor spreading and graph expansion*. CS and SE Seminar, Concordia Univ., Montreal, Canada, Mar 10 2014.

Anne-Marie Kermarrec. *Privacy-preserving distributed collaborative filtering*. Lecture at NICTA, Sydney, Australia, Jan 2014.

Anne-Marie Kermarrec. *Catch me if you can: Privacy-preserving micro-blogging*. Workshop on Distributed Systems, Univ. of Sydney, Jan 2014.

Anne-Marie Kermarrec. Scalable infrastructures for personalization. Morgenstern Colloquium, Inria Sophia-Antipolis, Apr 2014.

Anne-Marie Kermarrec. *Privacy-preserving distributed collaborative filtering*. Spring School METIS 2014, Marrakech, May 2014.

Anne-Marie Kermarrec. *Scalable Infrastructures for Personalization: from and ERC Grant to a startup*. Symposium on the Future of Cloud Computing, EIT-ICT Labs, Rennes, Jun 2014.

Anne-Marie Kermarrec. *Scalable infrastructures for personalization*. Yandex, Moscow, Jun 2014

Anne-Marie Kermarrec. *Scalable infrastructures for personalization*. MPI SWS Distinguihed Lecture, Saarbrucken, Nov 2014.

Michel Raynal. *What can be computed in a distributed system?* Workshop “From Programs to Systems: The Systems Perspective in Computing” in honor of Turing Award Winner Professor Joseph Sifakis, Apr 2014.

Michel Raynal. *From Turing to the clouds (on the computability power of distributed systems)*. 21th International Colloquium on Structural Information and Communication Complexity (SIROCCO), Jul 2014.

9.3. Teaching, supervision, juries

9.3.1. Teaching

Licence (bachelor) courses:

Simon Bouget, APF (Algorithmique et Programmation Fonctionnelle), 40 HEQTD (2 x 10 séances de 2 heures de TP), L1, Univ. Rennes 1 (ISTIC), France.

Davide Frey, Scalable Distributed Systems, 10 hours, M1, EIT/ICT Labs Master School, Univ. Rennes 1, France.

Davide Frey, Scalable Distributed Systems, 5 hours, M2, Univ. Rennes 1, France.

Michel Raynal, Introduction to distributed computing, 20h, M1, Univ. Rennes, France.

Michel Raynal, Distributed Computability, 20h, M2, Univ. Rennes, France.

François Taïani, Introduction to Operating Systems, 16h, 1st year of Engineering School (L3), ESIR / Univ. Rennes 1, France.

Master courses:

François Taïani, Synchronization and Parallel Programming, 48h, 2nd year of Engineering School (M1), ESIR / Univ. Rennes 1, France.

François Taïani, Distributed Systems, 48h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes 1, France.

François Taïani, Programming Technologies for the Cloud, 32h, M2, Univ. Rennes 1, France.

9.3.2. Supervision

PhD: Arnaud Jégou. *Harnessing the power of the implicit and explicit social networks through decentralization.* Sep 23 2014. Supervised by Davide Frey and Anne-Marie Kermarrec.

PhD in progress: Fabien André. *Calcul sur des données hétérogènes et modifiables dans les plateformes à large-échelle*. Started Nov 18 2013. Supervised by Anne-Marie Kermarrec (and Nicolas Le Scouarnec).

PhD in progress: Simon Bouget. *EMILIO: Emergent MIddleware for extra-Large-scale self adaptatIOn*. Started Sep 1 2014. Supervise by François Taïani

PhD in progress: Resmi Ariyattu Chandrasekharannair. *Towards decentralized federations for plug-based decentralized social networks*. Started Dec 1 2013. Supervised by François Taïani.

PhD in progress: Stéphane Delbruel. *Towards a decentralized embryomorphic storage System*. Started Oct 1 2013. Supervised by François Taïani and Davide Frey (since Oct 1 2014).

PhD in progress: Ali Gouta. *Caching and prefetching for efficient video services in mobile networks*. Started Jan 2 2012. Supervised by Anne-Marie Kermarrec (and Yannick Le Louédec).

PhD in progress: Nupur Mitall. *Infrastructure et algorithmes pour la recommandation de contenus cross application*. Started Dec 1 2013. Supervised by George Giakkoupis and Anne-Marie Kermarrec.

PhD in progress: Javier Olivares. *Web-Alter-Egos Platform: A peer-sourcing infrastructure for personalized privacy-aware event processing*. Started Sep 1 2013. Supervised by Anne-Marie Kermarrec.

PhD in progress: Antoine Rault. *Privacy through decentralization*. Started Oct 1 2012. Supervised by Davide Frey and Anne-Marie Kermarrec.

PhD in progress: Pierre-Louis Roman. *Epidemic distributed convergence for decentralized social networking*. Started Oct 1 2014. Supervised by Davide Frey and François Taïani.

PhD in progress: Julien Stainer. *Calculabilité distribuée*. Started Oct 1 2011. Supervised by Michel Raynal.

9.3.3. Juries

Anne-Marie Kermarrec was a member of the PhD juries for:

Ercan Ucan. *Data storage, transfers and communications in personal clouds*. ETH, Zurich, Feb 28 2014.

Oana Gog. *Matching user accounts across online social networks: Methods and application*. Univ. Paris 6, May 21 2014.

Abdelberi Chaabane. *Online social networks: Is it the end of privacy?*. Univ. Grenoble, May 22 2014.

Anne-Marie Kermarrec was a member of the HDR jury for Xavier Bonnaire. *Adaptabilité dans les systèmes répartis à large-échelle*. Univ. Paris 6, Jul 2014.

François Taïani was a member of the PhD juries for:

Bassem Debbabi. *Cube: A decentralised architecture-based framework for software self-management.* Univ. Grenoble (LIG), Jan 28 2014.

Adel Noureddine. *Towards a better understanding of the energy consumption of software systems.* Univ. Lille 1 (LIFL / Inria Lille), Mar 19 2014.

Romaric Ludinard. *Caractérisation locale de fautes dans les systèmes large échelle.* Univ. Rennes 1 (Inria Rennes - Bretagne Atlantique), Oct 2 2014.

François Taïani was a member of the HDR juries for:

Sébastien Gambs. *Respect de la vie privée dans la société de l'information.* Univ. Rennes 1, Jun 23 2014.

Yérom-David Bromberg. *Towards interoperable distributed systems* Univ. Bordeaux, Nov 28 2014.

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

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

[3] R. Friedman, A. Mostefaoui, S. Rajsbaum, M. Raynal. *Distributed agreement problems and their connection with error-correcting codes*, in "IEEE Transactions on Computers", 2007, vol. 56, n^{o} 7, pp. 865–875

[4] A. J. Ganesh, A.-M. Kermarrec, E. Le Merrer, L. Massoulié. *Peer counting and sampling in overlay networks based on random walks*, in "Distributed Computing", 2007, vol. 20, n^{o} 4, pp. 267-278

[5] 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

[6] 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

[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^{o} 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^{o} 1, pp. 125–129

Publications of the year

Doctoral Dissertations and Habilitation Theses

[10] A. JÉGOU. *Harnessing the power of implicit and explicit social networks through decentralization*, Université de Rennes 1, September 2014, https://hal.inria.fr/tel-01097143

Articles in International Peer-Reviewed Journals

[11] X. BAI, R. GUERRAOUI, A.-M. KERMARREC. *Personalizing Top-k Processing On-line in a Peer-to-Peer Social Tagging Network*, in "ACM Transactions on Internet Technology", 2014, vol. 13, n° 4, Article No. 11, https://hal.inria.fr/hal-00925990

[12] P. FRAIGNIAUD, G. GIAKKOUPIS. *Greedy routing in small-world networks with power-law degrees*, in "Distributed Computing", 2014, vol. 27, n° 4, pp. 231 - 253 [*DOI :* 10.1007/S00446-014-0210-Y], https://hal.inria.fr/hal-01097141

[13] A. GIURGIU, R. GUERRAOUI, K. HUGUENIN, A.-M. KERMARREC. *Computing in Social Networks*, in "Journal of Information and Computation", February 2014, vol. 234, pp. 3-16, SSS 2010 Special Issue [*DOI :* 10.1016/J.IC.2013.11.001], https://hal.archives-ouvertes.fr/hal-00827745

[14] R. GUERRAOUI, K. HUGUENIN, A.-M. KERMARREC, M. MONOD, S. PRUSTY, A. ROUMY. *Tracking Freeriders in Gossip-Based Content Dissemination Systems*, in "Computer Networks", May 2014, vol. 64, pp. 322-338 [*DOI :* 10.1016/J.COMNET.2014.02.023], https://hal.inria.fr/hal-00941107

[15] A.-M. KERMARREC. *Computer Science: Too Young to Fall into the Gender Gap*, in "IEEE Internet Computing", 2014, vol. 18, n° 3, https://hal.inria.fr/hal-01024738

[16] A.-M. KERMARREC, P. TRIANTAFILLOU. *XL Peer-to-Peer Pub/Sub Systems*, in "ACM Computing Surveys", 2014, vol. 46, n° 2, https://hal.inria.fr/hal-00853828

[17] F. TAÏANI, S. LIN, G. S. BLAIR. *GOSSIPKIT: A Unified Component Framework for Gossip*, in "IEEE Transactions on Software Engineering", March 2014, vol. 40, n° 2, pp. 123-136 [*DOI :* 10.1109/TSE.2013.50], https://hal.inria.fr/hal-01080198

International Conferences with Proceedings

[18] *Best Paper*
O. ALEXANDRA, A.-M. KERMARREC, K. ABERER. *Comparing the Predictive Capability of Social and Interest Affinity for Recommendations*, in "15th International Conference on Web Information System Engineering (WISE 2014)", Thessaloniki, Greece, 2014, https://hal.inria.fr/hal-01024730.

[19] F. ANDRÉ, A.-M. KERMARREC, E. LE MERRER, N. LE SCOUARNEC, G. STRAUB, A. VAN KEMPEN. *Archiving Cold Data in Warehouses with Clustered Network Coding*, in "EuroSys 2014", Amsterdam, Netherlands, ACM New York, NY, USA, April 2014 [*DOI :* 10.1145/2592798.2592816], https://hal.inria.fr/hal-00994660

[20] S. BOUGET, A.-M. KERMARREC, H. KERVADEC, F. TAÏANI. *Polystyrene: the Decentralized Data Shape That Never Dies*, in "The 2014 IEEE 34th International Conference on Distributed Computing Systems (ICDCS 2014)", Madrid, Spain, June 2014, pp. 288 - 297 [*DOI :* 10.1109/ICDCS.2014.37], https://hal.inria.fr/hal-01080608

[21] A. BOUTET, D. FREY, R. GUERRAOUI, A. JÉGOU, A.-M. KERMARREC. *Privacy-Preserving Distributed Collaborative Filtering*, in "NETYS", Marrakech, Morocco, May 2014, https://hal.inria.fr/hal-00975137

[22] A. BOUTET, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, R. PATRA. *HyRec: Leveraging Browsers for Scalable Recommenders*, in "Middleware 2014", Bordeaux, France, December 2014 [*DOI :* 10.1145/2663165.2663315], https://hal.inria.fr/hal-01080016

[23] N. CHILUKA, A.-M. KERMARREC, J. OLIVARES. *Scaling KNN Computation over Large Graphs on a PC*, in "Middleware 2014", Bourdeaux, France, December 2014 [*DOI :* 10.1145/2678508.2678513], https://hal.inria.fr/hal-01095557

[24] C. DELPORTE-GALLET, H. FAUCONNIER, M. RAYNAL. *An Exercise in Concurrency: From Non-blocking Objects to Fair Objects*, in "International Conference on Network-Based Information Systems(NBiS 2014)", Salerne, Italy, September 2014, https://hal.inria.fr/hal-01100781

[25] C. DELPORTE-GALLET, H. FAUCONNIER, M. RAYNAL. *Fair Synchronization in the Presence of Process Crashes and its Weakest Failure Detector*, in "33h Symposium on Reliable Distributed Systems (SRDS)", Nara, Japan, October 2014, pp. 161-170 [*DOI :* 10.1109/SRDS.2014.18], https://hal.inria.fr/hal-01100780

[26] D. FREY, M. GOESSENS, A.-M. KERMARREC. *Behave: Behavioral Cache for Web Content*, in "DAIS 2014", Berlin, Germany, LNCS, June 2014, pp. 89 - 103 [*DOI :* 10.1007/978-3-662-43352-2_8], https://hal.inria.fr/hal-01079976

[27] D. FREY, A.-M. KERMARREC, C. MADDOCK, A. MAUTHE, F. TAÏANI. *Adaptation for the Masses: Towards Decentralized Adaptation in Large-Scale P2P Recommenders*, in "Workshop on Adaptive and Reflective Middleware ARM 2014", Bordeaux, France, December 2014 [*DOI :* 10.1145/2677017.2677021], https://hal.inria.fr/hal-01080030

[28] G. GIAKKOUPIS. *Tight Bounds for Rumor Spreading with Vertex Expansion*, in "25th ACM-SIAM Symposium on Discrete Algorithms (SODA)", Portland, OR, United States, January 2014, pp. 801 - 815 [*DOI :* 10.1137/1.9781611973402.59], https://hal.inria.fr/hal-01097140

[29] G. GIAKKOUPIS, T. SAUERWALD, A. STAUFFER. *Randomized Rumor Spreading in Dynamic Graphs*, in "41st International Colloquium on Automata, Languages and Programming (ICALP)", Copenhagen, Denmark, July 2014, pp. 495 - 507 [*DOI :* 10.1007/978-3-662-43951-7_42], https://hal.inria.fr/hal-01097142

[30] G. GIAKKOUPIS, P. WOELFEL. *Randomized Mutual Exclusion with Constant Amortized RMR Complexity on the DSM*, in "55th IEEE Symposium on Foundations of Computer Science (FOCS)", Philadelphia, PA, United States, October 2014, https://hal.inria.fr/hal-01097144

[31] M. HERLIHY, S. RAJSBAUM, M. RAYNAL, J. STAINER. *Computing in the Presence of Concurrent Solo Executions*, in "LATIN 2014: Theoretical Informatics", Montevideo, Uruguay, 2014, vol. 8392, pp. 214-225 [*DOI :* 10.1007/978-3-642-54423-1_19], https://hal.inria.fr/hal-01097416

[32] D. IMBS, S. RAJSBAUM, M. RAYNAL, J. STAINER. *Reliable Shared Memory Abstraction on Top of Asynchronous Byzantine Message-Passing Systems*, in "Structural Information and Communication Complexity", Hida Takayama, Japan, Lecture Notes in Computer Science, Springer, 2014, vol. Volume 8576, pp. 37-53 [*DOI :* 10.1007/978-3-319-09620-9_5], https://hal.inria.fr/hal-01097383

[33] D. LIMA, H. MIRANDA, F. TAÏANI. *Towards a new model for cyber foraging*, in "ARM - The 13th Workshop on Adaptive and Reflective Middleware, in conjunction with ACM/IFIP/USENIX ACM International Middleware Conference", Bordeaux, France, December 2014 [*DOI :* 10.1145/2677017.2677023], https://hal.inria.fr/hal-01081147

[34] *Best Paper*
A. MOSTEFAOUI, M. HAMOUNA, M. RAYNAL. *Signature-Free Asynchronous Byzantine Consensus with $t < n/3$ and $O(n^2)$ Messages*, in "ACM PODC", Paris, France, Proceedings 34th Annual ACM Symposium on Principles of Distributed Computing, July 2014, pp. 2-9, https://hal.inria.fr/hal-00944019.

[35] J. PASTOR, M. BERTIER, F. DESPREZ, A. LÈBRE, F. QUESNEL, C. TEDESCHI. *Locality-aware Cooperation for VM Scheduling in Distributed Clouds*, in "Euro-Par 2014", Porto, Portugal, August 2014, https://hal.inria.fr/hal-00991530

[36] M. RAYNAL, J. STAINER, J. CAO, W. WU. *A Simple Broadcast Algorithm for Recurrent Dynamic Systems*, in "Advanced Information Networking and Applications (AINA)", Victoria, BC, Canada, 2014, pp. 933 - 939 [*DOI :* 10.1109/AINA.2014.115], https://hal.inria.fr/hal-01097426

[37] M. RAYNAL, J. STAINER, G. TAUBENFELD. *Brief announcement: distributed universality: contention-awareness; wait-freedom; object progress, and other properties*, in "Proceedings of the 2014 ACM symposium on Principles of distributed computing", Paris, France, 2014, pp. 206-208 [*DOI :* 10.1145/2611462.2611503], https://hal.inria.fr/hal-01097396

[38] M. RAYNAL, J. STAINER, G. TAUBENFELD. *Distributed Universality*, in "Principles of Distributed Systems", Cortina d'Ampezzo, Italy, 2014, vol. 8878, pp. 469-484 [*DOI :* 10.1007/978-3-319-14472-6_31], https://hal.inria.fr/hal-01097410

Conferences without Proceedings

[39] Z. AOUINI, T. MAMADOU, A. GOUTA, A.-M. KERMARREC, Y. LELOUEDEC. *Improving caching efficiency and quality of experience with CF-Dash*, in "NOSSDAV'14 - 24th ACM Workshop on Network and Operating Systems Support for Digital Audio and Video", Singapore, ACM, March 2014, 6 p. [*DOI :* 10.1145/2578260.2578268], https://hal.archives-ouvertes.fr/hal-01097325

Scientific Books (or Scientific Book chapters)

[40] A. LÈBRE, J. PASTOR, M. BERTIER, F. DESPREZ, J. ROUZAUD-CORNABAS, C. TEDESCHI, A.-C. ORGERIE, F. QUESNEL, G. FEDAK. *Beyond The Clouds, How Should Next Generation Utility Computing Infrastructures Be Designed?*, in "Cloud Computing: Challenges, Limitations and R&D Solutions", Z. MAHMOOD (editor), Springer, November 2014, https://hal.inria.fr/hal-01067888

Research Reports

[41] R. FRIEDMAN, M. RAYNAL, F. TAÏANI. *Fisheye Consistency: Keeping Data in Synch in a Georeplicated World*, IRISA, November 2014, n^o 2022, https://hal.inria.fr/hal-01081186

[42] G. GIAKKOUPIS, R. GUERRAOUI, A. JÉGOU, A.-M. KERMARREC, N. MITTAL. *Catch Me If You Can Privacy-Preserving Dissemination in Micro-Blogging*, May 2014, https://hal.inria.fr/hal-00993198

[43] D. IMBS, S. RAJSBAUM, M. RAYNAL, J. STAINER. *Reliable Shared Memory Abstractions on Top of Asynchronous t-Resilient Byzantine Message-passing Systems*, 2014, n^o PI-2018, https://hal.inria.fr/hal-00993400

[44] A.-M. KERMARREC, F. TAÏANI, J. M. TIRADO MARTIN. *Scaling Out Link Prediction with SNAPLE: 1 Billion Edges and Beyond*, Inria Rennes, January 2015, n^o RT-0454, https://hal.inria.fr/hal-01111459

[45] M. RAYNAL, J. STAINER, G. TAUBENFELD. *Distributed Universality: Contention-Awareness, Wait-freedom, Object Progress, and Other Properties*, January 2014, n^o PI-2014, https://hal.inria.fr/hal-00947657

Other Publications

[46] Z. BOUZID, A. MOSTEFAOUI, M. RAYNAL. *Minimal Synchrony for Asynchronous Byzantine Consensus*, January 2015, https://hal.inria.fr/hal-01103466

[47] R. CARVAJAL-GÓMEZ, D. FREY, M. SIMONIN, A.-M. KERMARREC. *WebGC: Browser-Based Gossiping [Live Demo/Poster]*, December 2014, Middleware 2014 [*DOI :* 10.1145/2678508.2678515], https://hal.inria.fr/hal-01080032

[48] A. MOSTEFAOUI, M. RAYNAL. *Communication and Agreement Abstractions in the Presence of Byzantine Processes*, February 2014, https://hal.inria.fr/hal-00942889

[49] A. MOSTEFAOUI, M. RAYNAL. *Asynchronous Byzantine Systems: From Multivalued to Binary Consensus with t < n/3, O(n 2) Messages, O(1) Time, and no Signature*, January 2015, https://hal.inria.fr/hal-01102496

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, pp. 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] 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

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

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

[57] 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

[58] 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

[59] 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

[60] 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