IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Rennes

Université Rennes 1

Activity Report 2015

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

Creation of the Project-Team: 2007 July 01

Keywords:

Computer Science and Digital Science:

1.1.1. - Multicore
1.1.6. - Cloud
1.1.7. - Peer to peer
1.1.9. - Fault tolerant systems
1.2.9. - Social Networks
1.3. - Distributed Systems
1.5.2. - Communicating systems
2.1.6. - Concurrent programming
2.1.7. - Distributed programming
2.6.2. - Middleware
3.1.3. - Distributed data
3.1.8. - Big data (production, storage, transfer)
3.5.1. - Analysis of large graphs
3.5.2. - Recommendation systems
4.8. - Privacy-enhancing technologies
7.1. - Parallel and distributed algorithms
7.9. - Graph theory

Other Research Topics and Application Domains:

6.3.1. - Web
6.3.3. - Network services
6.3.4. - Social Networks
6.4. - Internet of things
6.5. - Information systems
9.4.1. - Computer science
9.8. - Privacy

1. Members

Research Scientists

Anne-Marie Kermarrec [Team leader until March 2015, Inria, Senior Researcher, HdR]
Davide Frey [Inria, Researcher]
George Giakkoupis [Inria, Researcher]

Faculty Members

Francois Taiani [Team leader since April 2015, Univ. Rennes I, Professor, HdR]
Marin Bertier [INSA Rennes, Associate Professor]
David Bromberg [Univ. Rennes I, Professor, from Sep 2015, HdR]
Michel Raynal [Univ. Rennes I, Professor, HdR]

Engineers

Raziel Carvajal Gomez [Inria, until Sep 2015]
Jean Francois Verdonck [Inria, until Mar 2015, granted by FP7 EIT ICT LABS GA project]

PhD Students

Fabien Andre [Technicolor, granted by CIFRE]
Resmi Ariyattu Chandrasekharannair [Univ. Rennes I, granted by UEB]
Simon Bouget [Normale Sup Rennes]
Stephane Delbruel [Univ. Rennes I]
Eleni Kanellou [University of Crete, co-supervised, until Jun 2015]
Nupur Mittal [Inria, granted by Conseil Régional de Bretagne]
Javier Olivares [Universidad Técnica Federico Santa María, Valparaíso, Chile]
Antoine Rault [Inria]
Pierre-Louis Roman [Univ. Rennes I]
Olivier Ruas [Inria]
Julien Stainer [Univ. Rennes I, until May 2015]

Post-Doctoral Fellows

Zohir Bouzid [Univ. Rennes I, until Nov 2015]
Nitin Chiluka [Univ. Rennes I, until Jun 2015]
Hicham Lakhlef [Univ. Rennes I, from Sep 2015]
Juan Manuel Tirado Martin [Inria, until Mar 2015]

Visiting Scientists

Raluca Halalai [Université de Neuchâtel, until Sep 2015]
Hamouma Moumen [Université Béjaia, until Jun 2015]
Diogo Saraiva Lima [University of Lisbon, until Aug 2015]

Administrative Assistant

Virginie Desroches [Inria]

Others

Tom Ferragut [Inria, Student, from May 2015 until Jun 2015]
Nominoe Kervadec [Univ. Rennes I, Student, from Jul 2015 until Aug 2015]
Yasamin Nazari [Inria, Student, from Sep 2015]
Florent Gade [MEDIEGO, from Mar 2015]
Arnaud Jegou [MEDIEGO]
Xavier Lucas [MEDIEGO, until Oct 2015]
Lea Poisson [MEDIEGO, from May 2015]

2. Overall Objectives

2.1. 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. Scalability: Cloud computing meets p2p

Our second objective is to move forward in the area of **scalable infrastructures** for data intensive applications. In this context, we focus significant efforts on personalization systems, which represent 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.

2.3. 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.

2.4. 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?

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.

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.

2.5. 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. 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 [61], [56], [57]. 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.

3.2. Peer-to-peer overlay networks

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

A peer-to-peer overlay network logically connects peers on top of IP. Two main classes of such overlays dominate, structured and unstructured. The differences relate to the choice of the neighbors in the overlay, and the presence of an underlying naming structure. Overlay networks represent the main approach to build large-scale distributed systems that we retained. An overlay network forms a logical structure connecting participating entities on top of the physical network, be it IP or a wireless network. Such an overlay might form a structured overlay network [62], [63], [64] following a specific topology or an unstructured network [60], [65] where participating entities are connected in a random or pseudo-random fashion. In between, lie weakly structured peer-to-peer overlays where nodes are linked depending on a proximity measure providing more flexibility than structured overlays and better performance than fully unstructured ones. Proximity-aware overlays connect participating entities so that they are connected to close neighbors according to a given proximity metric reflecting some degree of affinity (computation, interest, etc.) between peers. We extensively use this approach to provide algorithmic foundations of large-scale dynamic systems.

3.3. Epidemic protocols

Epidemic algorithms, also called gossip-based algorithms [59], [58], 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 [58], monitoring, resource management, search, and more generally in building unstructured peer-to-peer networks.

3.4. 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.

3.5. 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. Highlights of the Year

4.1. Highlights of the Year

Anne-Marie Kermarrec created the Mediego Startup in April 2015

Michel Raynal was accepted as a new member of the Academia Europaea.

4.1.1. Awards

Fabien André and Anne-Marie Kermarrec received the Award "Prix du magazine la recherche" in Computer science for the Eurosys 2014 paper "Archiving cold data in warehouses with clustered network coding"[1].

5. New Software and Platforms

5.1. Brow2Brow

Browser-to-browser serverless toolboxes

FUNCTIONAL DESCRIPTION

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.

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

5.2. Dashboard

MediEgo Dashboard: A personalized news dashboard

KEYWORDS: Recommender system - Personalized stream of news - Dashboard

FUNCTIONAL DESCRIPTION

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

- Participants: Anne-Marie Kermarrec, Antoine Boutet, Yuri Barssi and Jean-Francois Verdonck
- Contact: Anne-Marie Kermarrec
- URL: http://www.mediego.com

5.3. GossipLib

KEYWORDS: Nat traversal - Epidemic protocols - Gossip protocols - Overlay maintenance - Peer-to-peer - Dissemination

FUNCTIONAL DESCRIPTION

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 as thus also provides the implementation of a number of standard gossip protocols that may be used out of the box or extended to build more complex protocols and applications. These include for example the peer-sampling protocols for overlay management.

GossipLib also provides facility for the configuration and deployment of applications as final-product but also as research prototype in environments like PlanetLab, clusters, network emulators, and even as event-based simulation. The code developed with GossipLib can be run both as a real application and in simulation simply by changing one line in a configuration file.

- Participants: Davide Frey, Ribeiro Heverson, Anne Marie Kermarrec, Imane Al Ifdal, and Ilham Ikbal
- Contact: Davide Frey
- URL: http://gossiplib.gforge.inria.fr/

5.4. MediEgo

KEYWORDS: Widget web - Social network - Recommendation
FUNCTIONAL DESCRIPTION

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.

- Participants: Antoine Boutet, Jacques Falcou, Jean-Francois Verdonck, Anne Marie Kermarrec, Sébastien Campion, Rachid Guerraoui, Davide Frey and Arnaud Jegou
- Partner: EPFL - Ecole Polytechnique Fédérale de Lausanne
- Contact: Sébastien Campion
- URL: http://www.mediego.com

5.5. WebGC

Web-based Gossip Communication
SCIENTIFIC DESCRIPTION

The library currently includes the implementation of two peer sampling protocols, Cyclon and the generic peer-sampling protocol from, as well as a clustering protocol. All protocols implement a common GossipProtocol “interface”
FUNCTIONAL DESCRIPTION

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.

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

5.6. WhatsUp

KEYWORD: Recommender system
FUNCTIONAL DESCRIPTION

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

- Participants: Davide Frey, Ribeiro Heverson, Antoine Boutet, Anne Marie Kermarrec, Arnaud Jegou, Rachid Guerraoui and Jean-Francois Verdonck
- Contact: Davide Frey

5.7. YALPS

KEYWORDS: Traffic-shaping - Nat traversal - Experimentation - Peer-to-peer - Simulator - Deployment

FUNCTIONAL DESCRIPTION

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 year's new work consisted in adding support for Non-Blocking I/O. It was carried out during the internship of Nominoe Kervadec.

- Participants: Davide Frey, Maxime Monod, Heverson Borba Ribeiro, Anne Marie Kermarrec and Arnaud Jegou, and Nominoe Kervadec
- Contact: Davide Frey
- URL: http://yalps.gforge.inria.fr/

5.8. p2p-allyours

Peer-to-Peer AllYours

FUNCTIONAL DESCRIPTION

P2P AllYours is customization of WhatsUp developed in the context of the EIT/ICT-Labs AllYours project. In addition to WhatsUp (the distributed recommender engine), p2p-AllYours comprises the following features: - a new web interface, which users can access through a local web-server integrated in P2PAllYours, - a set of automatic nodes (BOTs) that can extract news items from RSS feeds and insert them into the recommender system - a content-bootstrap that solves the issues related to bootstrapping the recommender system when a user connects for the first time. - An experiment management server that allows users to register for the application in the context of the testing program.

- Participants: Davide Frey, Heverson Borba Ribeiro, Raziel Carvajal Gomez, Arnaud Jegou and Anne-Marie Kermarrec
- Contact: Davide Frey

6. New Results

6.1. Models and Theory of Distributed Systems

6.1.1. Asynchronous Byzantine Systems: From Multivalued to Binary Consensus with t < n/3, O(n^2) Messages, O(1) Time, and no Signature

Participant: Michel Raynal.

This work [39] presents a new algorithm that reduces multivalued consensus to binary consensus in an asynchronous message-passing system made up of n processes where up to t may commit Byzantine failures. This algorithm has the following noteworthy properties: it assumes t < n/3 (and is consequently optimal from a resilience point of view), uses O(n^2) messages, has a constant time complexity, and does not use signatures. The design of this reduction algorithm relies on two new all-to-all communication abstractions. The first one allows the non-faulty processes to reduce the number of proposed values to c, where c is a small constant. The second communication abstraction allows each non-faulty process to compute a set of (proposed) values such that, if the set of a non-faulty process contains a single value, then this value belongs to the set of any non-faulty process. Both communication abstractions have an O(n^2) message complexity and a constant time complexity. The reduction of multivalued Byzantine consensus to binary Byzantine consensus is then a simple sequential use of these communication abstractions. To the best of our knowledge, this is the first asynchronous message-passing algorithm that reduces multivalued consensus to binary consensus with O(n^2) messages and constant time complexity (measured with the longest causal chain of messages) in the presence of up to t < n/3 Byzantine processes, and without using cryptography techniques. Moreover, this reduction algorithm tolerates message reordering by Byzantine processes.

This work was done in collaboration with Achour Mostefaoui from the LINA laboratory in Nantes.

6.1.2. Atomic Read/Write Memory in Signature-free Byzantine Asynchronous Message-passing Systems

Participant: Michel Raynal.

In this work [54] we designed a signature-free distributed algorithm which builds an atomic read/write shared memory on top of an n-process asynchronous message-passing system in which up to t < n/3 processes may commit Byzantine failures. From a conceptual point of view, this algorithm is designed to be as close as possible to the algorithm proposed by Attiya, Bar-Noy and Dolev (JACM 1995), which builds an atomic register in an n-process asynchronous message-passing system where up to t < n/2 processes may crash. The proposed algorithm is particularly simple. It does not use cryptography to cope with Byzantine processes, and is optimal from a t-resilience point of view (t < n/3). A read operation requires O(n) messages, and a write operation requires O(n 2) messages.

This work was done in collaboration with Achour Mostefaoui, Matoula Petrolia, and Claude Jard from the LINA laboratory in Nantes.

6.1.3. Intrusion-Tolerant Broadcast and Agreement Abstractions in the Presence of Byzantine Processes

Participant: Michel Raynal.

A process commits a Byzantine failure when its behavior does not comply with the algorithm it is assumed to execute. Considering asynchronous message-passing systems, this work [18] presents distributed abstractions, and associated algorithms, that allow non-faulty processes to correctly cooperate, despite the uncertainty created by the net effect of asynchrony and Byzantine failures. These abstractions are broadcast abstractions (namely, no-duplicity broadcast, reliable broadcast, and validated broadcast), and agreement abstraction (namely, consensus). While no-duplicity broadcast and reliable broadcast are well-known one-to-all communication abstractions, validated broadcast is a new all-to-all communication abstraction designed to address agreement problems. After having introduced these abstractions, we also presented an algorithm implementing validated broadcast on top of reliable broadcast. Then we presented two consensus algorithms, which are reductions of multivalued consensus to binary consensus. The first one is a generic algorithm, that can be instantiated with unreliable broadcast or no-duplicity broadcast, while the second is a consensus algorithm based on validated broadcast. Finally, a third algorithm is presented that solves the binary consensus problem. This algorithm is a randomized algorithm based on validated broadcast and a common coin. The presentation of all the abstractions and their algorithms is done incrementally. This work was done in collaboration with Achour Mostefaoui from the LINA laboratory in Nantes.

6.1.4. Minimal Synchrony for Asynchronous Byzantine Consensus

Participants: Michel Raynal, Zohir Bouzid.

Solving the consensus problem requires in one way or another that the underlying system satisfies some synchrony assumption. Considering an asynchronous message-passing system of n processes where (a) up to $t < n/3$ may commit Byzantine failures, and (b) each pair of processes is connected by two uni-directional channels (with possibly different timing properties), this work [24] investigates the synchrony assumption required to solve consensus, and presents a signature-free consensus algorithm whose synchrony requirement is the existence of a process that is an eventual t+1bisource. Such a process p is a correct process that eventually has (a) timely input channels from t correct processes and (b) timely output channels to t correct processes (these input and output channels can connect p to different subsets of processes). As this synchrony condition was shown to be necessary and sufficient in the stronger asynchronous system model (a) enriched with message authentication, and (b) where the channels are bidirectional and have the same timing properties in both directions, it follows that it is also necessary and sufficient in the weaker system model considered in this work. In addition to the fact that it closes a long-lasting problem related to Byzantine agreement, a noteworthy feature of the proposed algorithm lies in its design simplicity, which is a first-class property.

This work was done in collaboration with Achour Mostefaoui from the LINA laboratory in Nantes.

6.1.5. Signature-Free Asynchronous Binary Byzantine Consensus with t<n/3, O(n^2) Messages, and O(1) Expected Time

Participant: Michel Raynal.

This work [17] is on broadcast and agreement in asynchronous message-passing systems made up of n processes, and where up to t processes may have a Byzantine Behavior. Its first contribution is a powerful, yet simple, all-to-all broadcast communication abstraction suited to binary values. This abstraction, which copes with up to $t < n/3$ Byzantine processes, allows each process to broadcast a binary value, and obtain a set of values such that (1) no value broadcast only by Byzantine processes can belong to the set of a correct process, and (2) if the set obtained by a correct process contains a single value v, then the set obtained by any correct process contains v. The second contribution of this work is a new round-based asynchronous consensus algorithm that copes with up to $t < n/3$ Byzantine processes. This algorithm is based on the previous binary broadcast abstraction and a weak common coin. In addition of being signature-free and optimal with respect to the value of t, this consensus algorithm has several noteworthy properties: the expected number of rounds to decide is constant; each round is composed of a constant number of communication steps and involves $O(n^2)$ messages; each message is composed of a round number plus a constant number of bits. Moreover, the algorithm tolerates message reordering by the adversary (i.e., the Byzantine processes). This work was done in collaboration with Achour Mostefaoui from the LINA laboratory in Nantes, and Hamouma Moumen from Université de Béjaïa.

6.1.6. Specifying Concurrent Problems: Beyond Linearizability and up to Tasks

Participants: Michel Raynal, Zohir Bouzid.

Tasks and objects are two predominant ways of specifying distributed problems. A task specifies for each set of processes (which may run concurrently) the valid outputs of the processes. An object specifies the outputs the object may produce when it is accessed sequentially. Each one requires its own implementation notion, to tell when an execution satisfies the specification. For objects linearizability is commonly used, while for tasks implementation notions are less explored. Sequential specifications are very convenient, especially important is the locality property of linearizability, which states that linearizable objects compose for free into a linearizable object. However, most well-known tasks have no sequential specification. Also, tasks have no clear locality property. This work [26] introduces the notion of interval-sequential object. The corresponding implementation notion of interval-linearizability generalizes linearizability. Interval-linearizability allows to specify any task. However, there are sequential one-shot objects that cannot be expressed as tasks, under the simplest interpretation of a task. We also showed that a natural extension of the notion of a task is expressive enough to specify any interval-sequential object.

This work was done in collaboration with Armando Castaneda and Sergio Rajsbaumm from UNAM, Mexico.

6.1.7. *Test-and-Set in Optimal Space*

Participant: George Giakkoupis.

The test-and-set object is a fundamental synchronization primitive for shared memory systems. In [35] we address the number of registers (supporting atomic reads and writes) required to implement a one-shot test-and-set object in the standard asynchronous shared memory model with n processes. The best lower bound is $\log n - 1$ for obstruction-free and deadlock-free implementations, and recently a deterministic obstruction-free implementation using $O(\sqrt{n})$ registers was presented.

In [35] we close the gap between these existing upper and lower bounds by presenting a deterministic obstruction-free implementation of a one-shot test-and-set object from $\Theta(\log n)$ registers of size $\Theta(\log n)$ bits. Combining our obstruction-free algorithm with techniques from previous research, we also obtain a randomized wait-free test-and-set algorithm from $\Theta(\log n)$ registers, with expected step-complexity $\Theta(\log^* n)$ against the oblivious adversary. The core tool in our algorithm is the implementation of a deterministic obstruction-free *sifter* object, using only 6 registers. If k processes access a sifter, then when they have terminated, at least one and at most $\mathsf{floor}(2k+1)/3$ processes return “win” and all others return “lose”.

This is a joint work with Maryam Helmi (U. of Calgary), Lisa Higham (U. of Calgary), and Philipp Woelfel (U. of Calgary), supported by the RADCON Inria Associate Team.

6.2. Graph and Probabilistic Algorithms

6.2.1. *On the Quadratic Shortest Path Problem*

Participant: Davide Frey.

Finding the shortest path in a directed graph is one of the most important combinatorial optimization problems, having applications in a wide range of fields. In its basic version, however, the problem fails to represent situations in which the value of the objective function is determined not only by the choice of each single arc, but also by the combined presence of pairs of arcs in the solution. In this work [40] we model these situations as a Quadratic Shortest Path Problem, which calls for the minimization of a quadratic objective function subject to shortest-path constraints. We prove strong NP-hardness of the problem and analyze polynomially solvable special cases, obtained by restricting the distance of arc pairs in the graph that appear jointly in a quadratic monomial of the objective function. Based on this special case and problem structure, we devise fast lower bounding procedures for the general problem and show computationally that they clearly outperform other approaches proposed in the literature in terms of its strength.

6.2.2. *Tight Bounds on Vertex Connectivity Under Vertex Sampling*

Participant: George Giakkoupis.

A fundamental result by Karger (SODA 1994) states that for any λ-edge-connected graph with n nodes, independently sampling each edge with probability $p = \Omega(\log n/\lambda)$ results in a graph that has edge connectivity $\Omega(\lambda p)$, with high probability. In [27] we prove the analogous result for vertex connectivity, when sampling vertices. We show that for any k-vertex-connected graph G with n nodes, if each node is independently sampled with probability $p = \Omega(\sqrt{\log n/k})$, then the subgraph induced by the sampled nodes has vertex connectivity $\Omega(kp^2)$, with high probability. This bound improves upon the recent results of Censor-Hillel et al. (SODA 2014) and is existentially optimal.

This is a joint work with Keren Censor-Hillel (Technion), Mohsen Ghaffari (MIT), Bernhard Haeupler (Carnegie Mellon U.), and Fabian Kuhn (U. of Freiburg).

6.3. Scalable Systems

6.3.1. Being prepared in a sparse world: the case of KNN graph construction

Participants: Anne-Marie Kermarrec, Nupur Mittal, Francois Taiani.

This work presents KIFF [41], a generic, fast and scalable KNN graph construction algorithm. KIFF directly exploits the bipartite nature of most datasets to which KNN algorithms are applied. This simple but powerful strategy drastically limits the computational cost required to rapidly converge to an accurate KNN solution, especially for sparse datasets. Our evaluation on a representative range of datasets show that KIFF provides, on average, a speed-up factor of 14 against recent state-of-the art solutions while improving the quality of the KNN approximation by 18

This work was done in collaboration with Antoine Boutet from CNRS, Laboratoire Hubert Curien, Saint-Etienne, France.

6.3.2. Cheap and Cheerful: Trading Speed and Quality for Scalable Social Recommenders

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

Recommending appropriate content and users is a critical feature of on-line social networks. Computing accurate recommendations on very large datasets can however be particularly costly in terms of resources, even on modern parallel and distributed infrastructures. As a result, modern recommenders must generally trade-off quality and computational cost to reach a practical solution. This trade-off has however so far been largely left unexplored by the research community, making it difficult for practitioners to reach informed design decisions. In this work [37], we investigate to which extent the additional computing costs of advanced recommendation techniques based on supervised classifiers can be balanced by the gains they bring in terms of quality. In particular, we compare these recommenders against their unsupervised counterparts, which offer lightweight and highly scalable alternatives. We propose a thorough evaluation comparing 11 classifiers against 7 lightweight recommenders on a real Twitter dataset. Additionally, we explore data grouping as a method to reduce computational costs in a distributed setting while improving recommendation quality. We demonstrate how classifiers trained using data grouping can reduce their computing time by 6 while improving recommendations up to 22% when compared with lightweight solutions.

6.3.3. Fast Nearest Neighbor Search

Participants: Fabien André, Anne-Marie Kermarrec.

Nearest Neighbor (NN) search in high dimension is an important feature in many applications, such as multimedia databases, information retrieval or machine learning. Product Quantization (PQ) is a widely used solution which offers high performance, i.e., low response time while preserving a high accuracy. PQ represents high-dimensional vectors by compact codes. Large databases can therefore be stored in memory, allowing NN queries without resorting to slow I/O operations. PQ computes distances to neighbors using cache-resident lookup tables, thus its performance remains limited by (i) the many cache accesses that the algorithm requires, and (ii) its inability to leverage SIMD instructions available on modern CPUs.

To address these limitations, we designed a novel algorithm, PQ Fast Scan [19], that transforms the cache-resident lookup tables into small tables, sized to fit SIMD registers. This transformation allows (i) in-register lookups in place of cache accesses and (ii) an efficient SIMD implementation. PQ Fast Scan has the exact same accuracy as PQ, while having 4 to 6 times lower response time (e.g., for 25 million vectors, scan time is reduced from 74ms to 13ms).

This work was done in collaboration with Nicolas Le Scouarnec.

6.3.4. Holons: towards a systematic approach to composing systems of systems

Participants: Yérom-David Bromberg, François Taïani.

The world's computing infrastructure is increasingly differentiating into self-contained distributed systems with various purposes and capabilities (e.g. IoT installations, clouds, VANETs, WSNs, CDNs,. . .). Furthermore, such systems are increasingly being composed to generate systems of systems that offer value-added functionality. Today, however, system of systems composition is typically ad-hoc and fragile. It requires developers to possess an intimate knowledge of system internals and low-level interactions between their components. In this work [21], we outline a vision and set up a research agenda towards the generalised programmatic construction of distributed systems as compositions of other distributed systems. Our vision, in which we refer uniformly to systems and to compositions of systems as holons, employs code generation techniques and uses common abstractions, operations and mechanisms at all system levels to support uniform system of systems composition. We believe our holon approach could facilitate a step change in the convenience and correctness with which systems of systems can be built, and open unprecedented opportunities for the emergence of new and previously-unenvisaged distributed system deployments, analogous perhaps to the impact the mashup culture has had on the way we now build web applications.

This work was done in collaboration with Gordon Blair Geoff Coulson, and Yehia Elkhatib from Lancaster University (UK), Laurent Réveillère from University of Bordeaux / Labri, and Heverson Borba Ribeiro and Etienne Rivière from University of Neuchâtel (Switzerland).

6.3.5. Hybrid datacenter scheduling

Participant: Anne-Marie Kermarrec.

We address the problem of efficient scheduling of large clusters under high load and heterogeneous workloads. A heterogeneous workload typically consists of many short jobs and a small number of large jobs that consume the bulk of the cluster's resources.

Recent work advocates distributed scheduling to overcome the limitations of centralized schedulers for large clusters with many competing jobs. Such distributed schedulers are inherently scalable, but may make poor scheduling decisions because of limited visibility into the overall resource usage in the cluster. In particular, we demonstrate that under high load, short jobs can fare poorly with such a distributed scheduler.

We propose instead a new hybrid centralized/ distributed scheduler, called Hawk. In Hawk, long jobs are scheduled using a centralized scheduler, while short ones are scheduled in a fully distributed way. Moreover, a small portion of the cluster is reserved for the use of short jobs. In order to compensate for the occasional poor decisions made by the distributed scheduler, we propose a novel and efficient randomized work-stealing algorithm.

We evaluate Hawk using a trace-driven simulation and a prototype implementation in Spark. In particular, using a Google trace, we show that under high load, compared to the purely distributed Sparrow scheduler, Hawk improves the 50th and 90th percentile runtimes by 80% and 90% for short jobs and by 35% and 10% for long jobs, respectively. Measurements of a prototype implementation using Spark on a 100-node cluster confirm the results of the simulation. This work has been done in the context of the Inria/epfl research center and in collaboration with Pamela delgado, Florin Dinu and Willy Zwaenepoel from EPFL and published in Usenix ATC in 2015 [30].

6.3.6. Out-of-core KNN Computation

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

This work proposes a novel multi threading approach to compute KNN on large datasets by leveraging both disk and main memory efficiently. The main rationale of our approach is to minimize random accesses to disk, maximize sequential access to data and efficient usage of only a fraction of the available memory. This approach is evaluated by comparing its performance with a fully in-memory implementation of KNN, in terms of execution time and memory consumption. This multithreading approach outperforms the in-memory baseline in all cases when the large dataset does not fit in memory.

6.3.7. Scaling Out Link Prediction with SNAPLE

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

A growing number of organizations are seeking to analyze extra large graphs in a timely and resource-efficient manner. With some graphs containing well over a billion elements, these organizations are turning to distributed graph-computing platforms that can scale out easily in existing data-centers and clouds. Unfortunately such platforms usually impose programming models that can be ill suited to typical graph computations, fundamentally undermining their potential benefits. In this work [38], 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 very large graphs that a standard implementation of link prediction on GraphLab cannot handle. More precisely, we show that Snaple 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 standalone machine running a non-distributed state-of-the-art solution.

6.3.8. Similitude: Decentralised Adaptation in Large-Scale P2P Recommenders

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

Decentralised recommenders have been proposed to deliver privacy-preserving, personalised and highly scalable on-line recommendations. Current implementations tend, however, to rely on a hard-wired similarity metric that cannot adapt. This constitutes a strong limitation in the face of evolving needs. In this work [33], we propose a framework to develop dynamically adaptive decentralized recommendation systems. Our proposal supports a decentralised 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 mission.

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

6.3.9. Transactional Memory Recommenders

Participant: Anne-Marie Kermarrec.

The Transactional Memory (TM) paradigm promises to greatly simplify the development of concurrent applications. This led, over the years, to the creation of a plethora of TM implementations delivering wide ranges of performance across workloads. Yet, no universal TM implementation fits each and every workload. In fact, the best TM in a given workload can reveal to be disastrous for another one. This forces developers to face the complex task of tuning TM implementations, which significantly hampers the wide adoption of TMs. In this work, we address the challenge of automatically identifying the best TM implementation for a given workload. Our proposed system, ProteusTM, hides behind the TM interface a large library of implementations. Under the hood, it leverages an innovative, multi-dimensional online optimization scheme, combining two popular machine learning techniques: Collaborative Filtering and Bayesian Optimization. We integrated ProteusTM in GCC and demonstrated its ability to switch TM implementations and adapt several configuration parameters (e.g., number of threads). We extensively evaluated ProteusTM, obtaining average performance 3% less than the optimal, and gains up to 100 over static alternatives.

This work has been done in collaboration with Rachid Guerraoui from EPFL, Diego Didona Nuno Diegues, Ricardo Neves and Paolo Romano from INESC, Lisboa) and will be published in ASPLOS 2016 [31].

6.3.10. Want to scale in centralized systems? Think P2P

Participants: Anne-Marie Kermarrec, François Taïani.

Peer-to-peer (P2P) systems have been widely researched over the past decade, leading to highly scalable implementations for a wide range of distributed services and applications. A P2P system assigns symmetric roles to machines, which can act both as client and server. This distribution of responsibility alleviates the need for any central component to maintain a global knowledge of the system. Instead, each peer takes individual decisions based on a local and limited knowledge of the rest of the system, providing scalability by design. While P2P systems have been successfully applied to a wide range of distributed applications (multicast, routing, caches, storage, pub-sub, video streaming), with some highly visible successes (Skype, Bitcoin), they tend to have fallen out of fashion in favor of a much more cloud-centric vision of the current Internet. We think this is paradoxical, as cloud-based systems are themselves large-scale, highly distributed infrastructures. They reside within massive, densely interconnected datacenters, and must execute efficiently on an increasing number of machines, while dealing with growing volumes of data. Today even more than a decade ago, large-scale systems require scalable designs to deliver efficient services. In this work [16] we argue that the local nature of P2P systems is key for scalability regardless whether a system is eventually deployed on a single multi-core machine, distributed within a data center, or fully decentralized across multiple autonomous hosts. Our claim is backed by the observation that some of the most scalable services in use today have been heavily influenced by abstractions and rationales introduced in the context of P2P systems. Looking to the future, we argue that future large-scale systems could greatly benefit from fully decentralized strategies inspired from P2P systems. We illustrate the P2P legacy through several examples related to Cloud Computing and Big Data, and provide general guidelines to design large-scale systems according to a P2P philosophy.

6.3.11. 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 the final version of the library at WISE 2015 [53].

6.4. Privacy in User Centric Applications

6.4.1. Collaborative Filtering Under a Sybil Attack: Similarity Metrics do Matter!

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

Whether we are shopping for an interesting book or selecting a movie to watch, the chances are that a recommendation system will help us decide what we want. Recommendation systems collect information about our own preferences, compare them to those of other users, and provide us with suggestions on a variety of topics. But is the information gathered by a recommendation system safe from potential attackers, be them other users, or companies that access the recommendation system? And above all, can service providers protect this information while still providing effective recommendations? In this work, we analyze the effect of Sybil attacks on collaborative-filtering recommendation systems, and discuss the impact of different similarity metrics in the trade-off between recommendation quality and privacy. Our results, on a state-of-the-art recommendation framework and on real datasets show that existing similarity metrics exhibit a wide range of behaviors in the presence of Sybil attacks. Yet, they are all subject to the same trade off: Sybil resilience for recommendation quality. We therefore propose and evaluate a novel similarity metric that combines the best of both worlds: a low RMSE score with a prediction accuracy for Sybil users of only a few percent. A preliminary version of this work was published at EuroSec 2015 [32].

This work was done in collaboration with Antoine Boutet, and Rachid Guerraoui.

6.4.2. Decentralized view prediction for global content placement

Participants: Stéphane Delbruel, Davide Frey, François Taïani.

A large portion of today's Internet traffic originates from streaming and video services. Storing, indexing, and serving these videos is a daily engineering challenge that requires increasing amounts of efforts and infrastructures. One promising direction to improve video services consists in predicting at upload time where and when a new video might be viewed, thereby optimizing placement and caching decisions. Implementing such a prediction service in a scalable manner poses significant technical challenges. In this work [28], we address these challenges in the context of a decentralized storage system consisting of set-top boxes or end nodes. Specifically, we propose a novel data placement algorithm that exploits information about the tags associated with existing content, such as videos, and uses it to infer the number of views that newly uploaded content will have in each country.

6.4.3. Distance-Based Differential Privacy in Recommenders

Participant: Anne-Marie Kermarrec.

The upsurge in the number of web users over the last two decades has resulted in a significant growth of online information. This information growth calls for recommenders that personalize the information proposed to each individual user. Nevertheless, personalization also opens major privacy concerns. We designed D2P, a novel protocol that ensures a strong form of differential privacy, which we call distance-based differential privacy, and which is particularly well suited to recommenders. D2P avoids revealing exact user profiles by creating altered profiles where each item is replaced with another one at some distance. We evaluate D2P analytically and experimentally on MovieLens and Jester datasets and compare it with other private and non-private recommenders. This work has been done in the context of the Web-Alter-ego Google focused award and in collaboration with Rachid guerraoui, Rhicheek Patra and Masha Taziki from EPFL and published in PVLVB in 2015 [15].

6.4.4. Privacy-Conscious Information Diffusion in Social Networks

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

This work presents a distributed algorithm – Riposte [47], for information dissemination in social networks which guarantees to preserve the privacy of its users. RIPOSTE ensures that information spreads widely if and only if a large fraction of users find it interesting, and this is done in a "privacy-conscious" manner, namely without revealing the opinion of any individual user. Whenever an information item is received by a user, RIPOSTE decides to either forward the item to all the user's neighbors, or not to forward it to anyone. The decision is randomized and is based on the user's (private) opinion on the item, as well as on an upper bound s on the number of user's neighbors that have not received the item yet.

This work was done in collaboration with Rachid Guerraoui from EPFL, Switzerland.

6.4.5. Hide & Share: Landmark-based Similarity for Private knn Computation

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.

In this work [22], 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 Antoine Boutet from the University of St. Etienne, and with Jingjing Wang and Rachid Guerraoui from EPFL, Switzerland.

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 has been 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. 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. 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.

8.1.2. 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.

8.1.3. 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. International Initiatives

8.2.1. Inria International Labs

Anne-Marie Kermarrec has been scientific collaborator at EPFL, Lausanne, since February 2014.

Anne-Marie Kermarrec has been the scientific coordinator of the EPFL/Inria International Lab since February 2015.

Anne-Marie Kermarrec organized the *First EPFL/Inria Workshop*, Lausanne, January 2015.

8.2.2. Inria Associate Teams not involved in an Inria International Labs

8.2.2.1. RADCON

Title: Randomized Algorithms for Distributed Computing and Networks

International Partner (Institution - Laboratory - Researcher):

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

Start year: 2013

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. International Research Visitors

8.3.1. Visits of International Scientists

Shlomi Dolev, Jan 21-24

Frederic Mallmann-Trenn Feb 11-18, 2015

Emmanuel Godard April 23-24, 2015

Hamouma Moumen, June 1-30, 2015

Stevens Le Blond, July 14, 2015

Raluca Halalai Aug 1-15, 2015

Diogo Lima Aug 3-28, 2015

Damien Imbs, Oct 11-24, 2015

8.3.1.1. Internships

Tom Ferragut; May 18 to June 27. *Study and evaluation of effective recommendation algorithms* Supervised by Anne-Marie

Nominoe Kervadec; from June 1 to Aug 31 2015. *Non-blocking I/O in YALPS* Supervised by Davide Frey.

Yasamin Nazari. *Asynchronous vs. Synchronous Rumor Spreading* Sep 1 - Dec 15 2015 Supervised by George Giakkoupis.

8.3.2. Visits to International Teams

8.3.2.1. Research stays abroad

George Giakkoupis visited University of Calgary, Canada, 1-8 Mar, 18 Jun - 10 Jul, and 5-21 Dec, Simon Fraser University, Canada, 22 Feb - 1 Mar, and University of Cambridge, UK, 22-30 Jul.

8.3.2.2. Internships

Nupur Mittal did an internship at NICTA, Sydney from March 30, 2015 to July 1, 2015 under the supervision of Dr. Dali Kaafar. She also received Ecole Doctorale Mobility grant for the same.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. General chair, scientific chair

George Giakkoupis will be the General Chair of the 35th ACM Symposium on Principles of Distributed Computing (PODC 2016), Chicago, Illinois, USA, July 2016.

9.1.1.2. Member of the organizing committees

Davide Frey served as workshop and tutorial co-chair for the ACM/IFIP/USENIX International Conference on Middleware (Middleware).

George Giakkoupis served in the conference committee and the steering committee of the 34th ACM Symposium on Principles of Distributed Computing (PODC 2015), Donostia-San Sebastián, Spain, July 2015.

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

Anne-Marie Kermarrec has been a member of the Inria Parity committee (and leader of the recrutment group) since September 2015.

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.

François Taïani has been a member of the Steering Committee of *the annual ACM/IFIP/USENIX Middleware conference* since 2014.

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

9.1.2. Scientific events selection

9.1.2.1. Member of the conference program committees

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

14th IEEE International Conference on Peer-to-Peer Computing (P2P), Cambridge, MA, USA, Sep 2015.

ACM/IFIP/USENIX International Conference on Middleware (Middleware), Vancouver, Canada, Dec 2015.

High Performance Computing (HiPC) 2015, Bangalore, India, Dec 2015.

George Giakkoupis served in the program committee of the 17th International Conference on Distributed Computing and Networking (ICDCN 2016), Singapore, Jan 2016.

Anne-Marie Kermarrec was a program committee member of the conferences.

Middleware 2015, Vancouver, BC, Canada, December 2015.

Eurosys 2015, Bordeaux, France, April 2015.

SRDS 2015, Montreal, Canada, October 2015.

Systor 2015, Haifa, Israel, May 2015.

Netys 2015, Agadir, Moroco May 2015.

François Taïani was a program committee member of the conferences.

Middleware 2015, Vancouver, BC, Canada, December 2015.

IEEE International Conference on Peer-to-Peer Computing (P2P), Cambridge, MA, USA, Sep. 2015.

IEEE 34th Symposium on Reliable Distributed Systems (SRDS 2015), Montreal, Canada, October 2015.

The 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2015), Rio do Janeiro, Brazil, June 2015.

ICDCS 2015, The 35th IEEE International Conference on Distributed Computing Systems (ICDCS 2015), Colombus, Ohio, USA, June-July 2015.

9.1.3. Journal

9.1.3.1. Member of the editorial boards

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

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

9.1.4. Invited talks

George Giakkoupis. *Vertex connectivity under sampling.* Seminar of the Computer Science and Telecommunications Department, Ecole Normale Supérieure, Rennes, France, Sep 15 2015.

George Giakkoupis. *Bounds on the voter model in terms of graph expansion.* Random Walks on Random Graphs and Applications Worshop, Eurandom, TU Eindhoven, The Netherlands, Apr 14 2015.

George Giakkoupis. *Rumor spreading: Techniques and results*. Tutorial, 8th Winter School on Hot Topics in Distributed Computing (HTDC), Flaine, France, Mar 18 2015.

George Giakkoupis. *Vertex connectivity under vertex sampling*. PIMS Algorithmic Theory of Networks Seminar, University of Calgary, Canada, Mar 6 2015.

Anne-Marie Kermarrec, *TM Recommenders*, Invited speaker at the 7th workshop on the Theory of Transactional Memory, Donostia, Spain 2015.

Anne-Marie Kermarrec, *Scalable personalization infrastructures*, IST Austria, April 2015.

Anne-Marie Kermarrec, *TM Recommenders*, Rennes, November 2015

Nupur Mittal, *Privacy-Conscious Information Diffusion in Social Networks*, Rennes, November 2015.

Michel Raynal, *Concurrent Systems: Hybrid Object Implementations and Abortable Objects.* at Euro-Par 2015.

Michel Raynal, *Communication Patterns and Input Patterns in Distributed Computing* at SIROCCO 2015.

François Taïani. *Engineering Scale: Software and Distribution for Tomorrow's World*. Keynote presentation at the Journées Nationales 2015 of the GDR GPL (France's research federation on software engineering and programming), Bordeaux, France, 9-12 June 2015.

François Taïani. *Large-Scale Systems And Self-Organization*. Invited lecture at the Doctoral School "Interoperability and automatic adaptation for novel generation middleware systems" of Doctoral Program in Computer Science of CUSO (Conférence Universitaire de Suisse Occidentale), Neuchâtel, Switzerland, November 26th - 27th, 2015.

9.1.5. Leadership within the scientific community

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

Anne-Marie Kermarrec co-organized the *Workshop Inria/Technicolor on workshop on storage and data analytics* , Rennes, Nov. 2015.

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

Anne-Marie Kermarrec was a member of the Inria scientific board (Bureau du comité des projets) in Rennes until July 2015.

François Taïani was the chair of the hiring committee for the position of *Professor at University of Rennes 1 - ESIR*.

9.1.6. Research administration

Davide Frey is scientific correspondent for the DPEI in Rennes.

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

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

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

Licence : Simon Bouget, SYS1 (Introduction aux Systèmes d'exploitation), 10 HEQTD (5 séances de 2 heures de TP), ESIR1 (L3), Université de Rennes 1 (ESIR), France.

Master : Simon Bouget, TLC (Techniques Logicielles pour le Cloud computing), 14 HEQTD (7 séances de 2 heures de TP), M2 Génie Logiciel Alternance, Université de Rennes 1 (ISTIC), France.

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

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

Engineering School: François Taïani, Distributed Systems, 24h, 3rd year of Engineering School (M2), ESIR / Université of Rennes 1, France.

Engineering School: François Taïani, Parallel Algorithms for Big Data, 24h, 3rd year of Engineering School (M2), ESIR / Université of Rennes 1, France.

Engineering School: François Taïani, Introduction to Operating Systems, 24h, 1st year of Engineering School (L3), ESIR / Université of Rennes 1, France.

Master: François Taïani, Programming Technologies for the Cloud, 32h, M2, Université of Rennes 1, France.

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

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

9.2.2. Supervision

PhD: Ali Gouta *Caching and prefetching for efficient video services in mobile networks* [11]. Université de Rennes 1, January 15, 2015. Supervised by Anne-Marie Kermarrec.

PhD: Julien Stainer. *Computability Abstractions for Fault-tolerant Asynchronous Distributed Computing* [13]. Université de Rennes 1, March 18, 2015. Supervised by Michel Raynal.

PhD: Eleni Kanellou. *Data Structures for Current Multi-core and Future Many-core architectures* [12]. Université de Rennes 1, December 14, 2015. Supervised by Michel Raynal.

PhD in progress : Antoine RAULT, *User privacy in collaborative filtering systems*. Started on 1st October 2012. Supervised by Anne-Marie Kermarrec, Davide Frey.

PhD in progress : Stéphane Delbruel, Towards a Decentralized Embryomorphic Storage System, 1st October 2013, François Taïani and Davide Frey (since 1st October 2014).

PhD in progress : Resmi Ariyattu Chandrasekharannair, Towards Decentralized Federations for Plug-based Decentralized Social Networks, 1st December 2013, François Taïani.

PhD in progress : Simon Bouget, EMILIO: Emergent MIddleware for extra-Large-scale self adaptatIOn, 1st September 2014, François Taïani.

PhD in progress : Pierre-Louis Roman, Epidemic Distributed Convergence for Decentralized Social Networking, 1st October 2014, François Taïani, Davide Frey.

PhD in progress: Javier Olivares, A peer-sourcing infrastructure for personalized privacy-aware event processing, 1st October 2013, Anne-Marie Kermarrec.

PhD in progress : Olivier Ruas, Dynamic learning and recommendations in very large distributed com-puting infrastructures, 1st October 2015, Anne-Marie Kermarrec, François Taïani.

PhD in progress: Nupur Mittal. *Infrastructure et algorithmes pour la recommandation de contenus cross application*, December 1 2013, George Giakkoupis and Anne-Marie Kermarrec.

9.2.3. Juries

Anne-Marie Kermarrec was in the HDR juries of Vincent Gramoli (June 2015) and Sebastien Monnet (November 2015) (LIP 6).

François Taïani was a reviewer of the PhD for Raluca Diaconu. *Scalability for Virtual Worlds*. Université Pierre et Marie Curie, January 23, 2015.

François Taïani was the chair of the PhD jury for Julien Stainer. *Computability Abstractions for Fault-tolerant Asynchronous Distributed Computing*, Université de Rennes 1, March 18, 2015.

François Taïani was a reviewer of the PhD for Camille Fayollas. *Architecture logicielle générique et approche à base de modèles pour la sûreté de fonctionnement des systèmes interactifs critiques*. Université Toulouse 3 Paul Sabatier, July 21, 2015.

François Taïani was a reviewer of the PhD for Pierre André. *Test de systèmes ubiquitaires avec prise en compte explicite de la mobilité*. Université Toulouse 3 Paul Sabatier, November 17, 2015.

François Taïani was the chair of the PhD jury for Erwan Bousse. *Execution Trace Management to Support Dynamic V&V for Executable DSMLs*. Université de Rennes 1, December 3, 2015.

François Taïani was the chair of the PhD jury for Eleni Kanellou. *Data Structures for Current Multi-core and Future Many-core architectures*. Université de Rennes 1, December 14, 2015.

François Taïani served as external examiner on the PhD committee of Xi Li. *Workload Consolidation in Large-scale Data Centers*. University College Dublin (UCD), Ireland, December 16, 2015.

François Taïani was a member of the HDR jury for Benoît Parrein. *Le code à effacement Mojette : Applications dans les réseaux et dans le Cloud*. Université de Nantes, June 16, 2015.

François Taïani was a reviewer of the HDR of Gerson Sunyé. *A Model-Based Approach for Testing Large Scale Systems*. Université de Nantes, November 10, 2015.

9.3. Popularization

Anne-Marie Kermarrec wrote three articles on women in computer science: on the blog of the SIF Binaire in March 2015.

10. Bibliography

Major publications by the team in recent years

[1] 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

[2] 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

[3] 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

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

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

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

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

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

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

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

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. GOUTA. *Caching and prefetching for efficient video services in mobile networks*, University of Rennes 1, January 2015, https://hal.inria.fr/tel-01256966

[12] E. KANELLOU. *Data Structures for Current Multi-core and Future Many-core Architectures*, Université de Rennes 1, December 2015, https://hal.inria.fr/tel-01256954

[13] J. STAINER. *Computability Abstractions for Fault-tolerant Asynchronous Distributed Computing*, Université Rennes 1, March 2015, https://hal.inria.fr/tel-01256926

Articles in International Peer-Reviewed Journals

[14] A. BOUTET, D. FREY, R. GUERRAOUI, A. JÉGOU, A.-M. KERMARREC. *Privacy-Preserving Distributed Collaborative Filtering*, in "Computing", 2015, https://hal.inria.fr/hal-01251314

[15] R. GUERRAOUI, A.-M. KERMARREC, R. PATRA, M. TAZIKI. *D2P: Distance-Based Differential Privacy in Recommenders*, in "Proceedings of the VLDB Endowment", 2015, vol. 8, n^o 8, pp. 862-873 [*DOI :* 10.14778/2757807.2757811], https://hal.inria.fr/hal-01183859

[16] A.-M. KERMARREC, F. TAÏANI. *Want to scale in centralized systems? Think P2P*, in "Journal of Internet Services and Applications", August 2015, 18 p. [*DOI :* 10.1186/s13174-015-0029-1], https://hal.inria.fr/hal-01199734

[17] A. MOSTÉFAOUI, H. MOUMEN, M. RAYNAL. *Signature-Free Asynchronous Binary Byzantine Consensus with t<n/3, O(n²) Messages, and O(1) Expected Time*, in "Journal of the ACM (JACM)", December 2015, vol. 62, pp. 1000-1020 [*DOI :* 10.1145/2785953], https://hal.archives-ouvertes.fr/hal-01176110

[18] A. MOSTÉFAOUI, M. RAYNAL. *Intrusion-Tolerant Broadcast and Agreement Abstractions in the Presence of Byzantine Processes*, in "IEEE Transactions on Parallel and Distributed Systems", 2015, pp. 1-24 [*DOI :* 10.1109/TPDS.2015.2427797], https://hal.archives-ouvertes.fr/hal-01176113

International Conferences with Proceedings

[19] F. ANDRÉ, A.-M. KERMARREC, N. LE SCOUARNEC. *Cache locality is not enough: High-Performance Nearest Neighbor Search with Product Quantization Fast Scan*, in "42nd International Conference on Very Large Data Bases", New Delhi, India, September 2016, vol. 9, n^o 4, 12 p. , https://hal.inria.fr/hal-01239055

[20] R. ARIYATTU, F. TAÏANI. *Fluidify: Decentralized Overlay Deployment in a Multi-Cloud World*, in "DAIS", Inria Grenoble, France, June 2015, 14 p. , https://hal.archives-ouvertes.fr/hal-01168137

[21] G. BLAIR, Y.-D. BROMBERG, G. COULSON, Y. ELKHATIB, L. RÉVEILLÈRE, H. BORBA RIBEIRO, E. RIVIÈRE, F. TAÏANI. *Holons: towards a systematic approach to composing systems of systems*, in "The 14th International Workshop on Adaptive and Reflective Middleware (ARM 2015)", Vancouver, France, December 2015 [*DOI :* 10.1145/2834965.2834970], https://hal.inria.fr/hal-01245251

[22] A. BOUTET, D. FREY, R. GUERRAOUI, A.-M. KERMARREC, A. RAULT, F. TAÏANI, J. WANG. *Hide & Share: Landmark-based Similarity for Private KNN Computation*, in "45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)", Rio de Janeiro, Brazil, June 2015, pp. 263-274 [*DOI :* 10.1109/DSN.2015.60], https://hal.archives-ouvertes.fr/hal-01171492

[23] A. BOUTET, A.-M. KERMARREC, N. MITTAL, F. TAÏANI. *Being prepared in a sparse world: the case of KNN graph construction*, in "International Conference on Data Engineering (ICDE)", Helsinki, Finland, May 2016, https://hal.inria.fr/hal-01251010

[24] Z. BOUZID, A. MOSTÉFAOUI, M. RAYNAL. *Minimal Synchrony for Byzantine Consensus*, in "Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing (PODC 2015)", Donostia-San Sebastián, Spain, July 2015, pp. 461-470 [*DOI :* 10.1145/2767386.2767418], https://hal.archives-ouvertes.fr/hal-01178651

[25] Z. BOUZID, M. RAYNAL, P. SUTRA. *Brief Announcement: Anonymous Obstruction-free (n, k)-Set Agreement with n-k+1 Atomic Read/Write Registers*, in "DISC 2015", Tokyo, Japan, Y. MOSES, M. ROY (editors), 29th International Symposium on Distributed Computing, Springer-Verlag Berlin Heidelberg, October 2015, vol. LNCS 9363, https://hal.archives-ouvertes.fr/hal-01207902

[26] A. CASTAÑEDA, S. RAJSBAUM, M. RAYNAL. *Specifying Concurrent Problems: Beyond Linearizability and up to Tasks*, in "DISC 2015", Tokyo, Japan, Y. MOSES, M. ROY (editors), 29th International Symposium on Distributed Computing, Springer-Verlag Berlin Heidelberg, October 2015, vol. LNCS 9363 [*DOI :* 10.1007/978-3-662-48653-5_28], https://hal.archives-ouvertes.fr/hal-01207140

[27] K. CENSOR-HILLEL, M. GHAFFARI, G. GIAKKOUPIS, B. HAEUPLER, F. KUHN. *Tight Bounds on Vertex Connectivity Under Vertex Sampling*, in "26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA 2015)", San Diego, CA, United States, January 2015, pp. 2006-1018 [*DOI :* 10.1137/1.9781611973730.133], https://hal.inria.fr/hal-01250519

[28] S. DELBRUEL, D. FREY, F. TAÏANI. *Decentralized view prediction for global content placement*, in "Middleware 2015 : ARM Workshop", Vancouver, Canada, December 2015 [*DOI :* 10.1145/2834965.2834974], https://hal.inria.fr/hal-01247159

[29] S. DELBRUEL, F. TAÏANI. *Géodistribution des tags et des vues dans Youtube*, in "Conférence d'informatique en Parallélisme, Architecture et Système. Compas'2015", Lille, France, June 2015, https://hal.inria.fr/hal-01162568

[30] P. DELGADO, F. DINU, A.-M. KERMARREC, W. ZWAENEPOEL. *Hawk: Hybrid Datacenter Scheduling*, in "2015 USENIX Annual Technical Conference", Santa-Clara, United States, USENIX Association, July 2015, pp. 499-510, https://hal.inria.fr/hal-01183857

[31] D. DIDONA, N. DIEGUES, R. GUERRAOUI, A.-M. KERMARREC, R. NEVES, P. ROMANO. *ProteusTM: Abstraction Meets Performance in Transactional Memory*, in "Twenty First International Conference on Architectural Support for Programming Languages and Operating Systems", Atlanta, United States, April 2016, https://hal.inria.fr/hal-01250459

[32] D. FREY, R. GUERRAOUI, A.-M. KERMARREC, A. RAULT. *Collaborative Filtering Under a Sybil Attack: Analysis of a Privacy Threat*, in "Eighth European Workshop on System Security EuroSec 2015", Bordeaux, France, April 2015 [*DOI :* 10.1145/2751323.2751328], https://hal.inria.fr/hal-01158723

[33] D. FREY, A.-M. KERMARREC, C. MADDOCK, A. MAUTHE, P.-L. ROMAN, F. TAÏANI. *Similitude: Decentralised Adaptation in Large-Scale P2P Recommenders*, in "The 15th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS)", Grenoble, France, June 2015 [*DOI :* 10.1007/978-3-319-19129-4_5], https://hal.inria.fr/hal-01138365

[34] G. GIAKKOUPIS, R. GUERRAOUI, A. JÉGOU, A.-M. KERMARREC, N. MITTAL. *Privacy-Conscious Information Diffusion in Social Networks*, in "DISC 2015", Tokyo, Japan, Y. MOSES, M. ROY (editors), 29th International Symposium on Distributed Computing, Springer-Verlag Berlin Heidelberg, October 2015, vol. LNCS 9363 [*DOI :* 10.1007/978-3-662-48653-5_32], https://hal.archives-ouvertes.fr/hal-01207162

[35] G. GIAKKOUPIS, M. HELMI, L. HIGHAM, P. WOELFEL. *Test-and-Set in Optimal Space*, in "47th Annual ACM Symposium on Theory of Computing (STOC 2015)", Portland, OR, United States, June 2015, pp. 615-623 [*DOI :* 10.1145/2746539.2746627], https://hal.inria.fr/hal-01250520

[36] A. GOUTA, D. HAUSHEER, A.-M. KERMARREC, C. KOCH, Y. LELOUEDEC, J. RUCKERT. *CPSys: A system for mobile video prefetching*, in "IEEE 23rd International Symposium on Modelling Analysis and Simulation of Computer and Telecommunication Systems - MASCOTS 2015", Atlanta, United States, IEEE, October 2015, https://hal.inria.fr/hal-01183855

[37] A.-M. KERMARREC, F. TAÏANI, J. M. TIRADO MARTIN. *Cheap and Cheerful: Trading Speed and Quality for Scalable Social Recommenders*, in "Proceedings of the 15th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS-2015)", Grenoble, France, Springer International Publishing, June 2015, 14 p. [*DOI :* 10.1007/978-3-319-19129-4_11], https://hal.inria.fr/hal-01170757

[38] A.-M. KERMARREC, F. TAÏANI, J. M. TIRADO MARTIN. *Scaling Out Link Prediction with SNAPLE*, in "16th Annual ACM/IFIP/USENIX Middleware Conference", Vancouver, Canada, December 2015, 12 p. [*DOI :* 10.1145/2814576.2814810], https://hal.inria.fr/hal-01244663

[39] A. MOSTÉFAOUI, M. RAYNAL. *Asynchronous Byzantine Systems: From Multivalued to Binary Consensus with t < n/3, O(n^2) Messages, O(1) Time, and no Signature*, in "22nd International Colloquium on Structural Information & Communication Complexity (SIROCCO'15)", Montserrat, Spain, July 2015, https://hal.archives-ouvertes.fr/hal-01176117

[40] B. ROSTAMI, F. MALUCELLI, D. FREY, C. BUCHHEIM. *On the Quadratic Shortest Path Problem*, in "14th International Symposium on Experimental Algorithms", Paris, France, June 2015 [*DOI :* 10.1007/978-3-319-20086-6_29], https://hal.inria.fr/hal-01251438

Research Reports

[41] A. BOUTET, A.-M. KERMARREC, N. MITTAL, F. TAÏANI. *Impressively fast and efficient KNN construction*, University Saint Etienne, June 2015, https://hal.inria.fr/hal-01169782

[42] Z. BOUZID, M. RAYNAL, P. SUTRA. *Anonymous Obstruction-free (n, k)-Set Agreement with $n - k + 1$ Atomic Read/Write Registers*, univzrité de rennes 1, 2015, n° 2027, 18 p. , https://hal.inria.fr/hal-01169693

[43] S. DELBRUEL, D. FREY, F. TAÏANI. *Exploring the Geography of Tags in Youtube Views*, IRISA, Inria Rennes ; Inria, April 2015, n° RT-0461, 27 p. , https://hal.inria.fr/hal-01157867

[44] D. FREY, R. FRIEDMAN, A. MOSTEFAOUI, M. PERRIN, M. RAYNAL, F. TAÏANI. *D.1.2 – Modular quasi-causal data structures*, LINA-University of Nantes ; IRISA, November 2015, n° D1.2, https://hal.archives-ouvertes.fr/hal-01223119

[45] D. FREY, H. LAKHLEF, M. RAYNAL. *Optimal Collision/Conflict-free Distance-2 Coloring in Synchronous Broadcast/Receive Tree Networks*, IRISA, Inria Rennes, December 2015, n° 2030, 19 p. , https://hal.inria.fr/hal-01248428

[46] D. FREY, A. MOSTEFAOUI, M. PERRIN, F. TAÏANI. *D.1.1 – Survey on Weak Consistency Approaches for Large-Scale Systems*, LINA-University of Nantes ; IRISA, June 2015, n° D1.1, https://hal.archives-ouvertes.fr/hal-01174203

[47] G. GIAKKOUPIS, R. GUERRAOUI, A. JÉGOU, A.-M. KERMARREC, N. MITTAL. *Privacy-Conscious Information Diffusion in Social Networks*, Inria Rennes - Bretagne Atlantique, August 2015 [*DOI :* 10.1007/978-3-662-48653-5_32], https://hal.archives-ouvertes.fr/hal-01184246

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

[49] A. LEBRE, J. PASTOR, . THE DISCOVERY CONSORTIUM. *The DISCOVERY Initiative - Overcoming Major Limitations of Traditional Server-Centric Clouds by Operating Massively Distributed IaaS Facilities*, Inria, September 2015, n° RR-8779, 14 p. , https://hal.inria.fr/hal-01203648

[50] A. MOSTEFAOUI, M. RAYNAL. *Time-Efficient Read/Write Register in Crash-prone Asynchronous Message-Passing Systems*, IRISA, January 2016, 14 p. , https://hal.inria.fr/hal-01256067

[51] B. NÉDELEC, J. TANKE, D. FREY, P. MOLLI, A. MOSTEFAOUI. *Spray: an Adaptive Random Peer Sampling Protocol*, LINA-University of Nantes ; Inria Rennes - Bretagne Atlantique, September 2015, https://hal.archives-ouvertes.fr/hal-01203363

Other Publications

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

[53] R. CARVAJAL-GÓMEZ, D. FREY, M. SIMONIN, A.-M. KERMARREC. *WebGC Gossiping on Browsers without a Server [Live Demo/Poster]*, November 2015, Web Information System Engineering, Poster, https://hal.inria.fr/hal-01251787

[54] A. MOSTEFAOUI, M. PETROLIA, M. RAYNAL, C. JARD. *Atomic Read/Write Memory in Signature-free Byzantine Asynchronous Message-passing Systems*, December 2015, working paper or preprint, https://hal.inria.fr/hal-01238765

[55] 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, working paper or preprint, https://hal.inria.fr/hal-01102496

References in notes

[56] 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

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

[58] 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

[59] 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

[60] 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

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

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

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

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

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