IN PARTNERSHIP WITH:
Institut national des sciences appliquées de Rennes

Université Rennes 1

Activity Report 2017

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, end of the Project-Team: 2017 December 31

Keywords:

Computer Science and Digital Science:

A1.1.1. - Multicore, Manycore
A1.1.6. - Cloud
A1.1.7. - Peer to peer
A1.1.9. - Fault tolerant systems
A1.2.9. - Social Networks
A1.3. - Distributed Systems
A1.5.2. - Communicating systems
A2.1.6. - Concurrent programming
A2.1.7. - Distributed programming
A2.6.2. - Middleware
A3.1.3. - Distributed data
A3.1.8. - Big data (production, storage, transfer)
A3.5.1. - Analysis of large graphs
A3.5.2. - Recommendation systems
A4.8. - Privacy-enhancing technologies
A7.1. - Algorithms
A8.7. - Graph theory

Other Research Topics and Application Domains:

B6.3.1. - Web
B6.3.3. - Network Management
B6.3.4. - Social Networks
B6.4. - Internet of things
B6.5. - Information systems
B9.4.1. - Computer science
B9.8. - Privacy

1. Personnel

Research Scientists

Davide Frey [Inria, Researcher]
George Giakkoupis [Inria, Researcher]
Anne-Marie Kermarrec [Inria, Senior Researcher, until Jan 2017, HDR]

Faculty Members

Francois Taiani [Team leader, Univ de Rennes I, Professor, HDR]
Marin Bertier [INSA Rennes, Associate Professor]
David Bromberg [Univ de Rennes I, Professor, HDR]
Michel Raynal [Univ de Rennes I, Professor, HDR]

Post-Doctoral Fellow

Zohir Bouzid [Univ de Rennes I, until May 2017]

PhD Students

Simon Bouget [Univ de Rennes I, until Oct 2017]
Louison Gitzinger [Univ de Rennes I, from Oct 2017]
Adrien Luxey [Univ de Rennes I]
Pierre-Louis Roman [Univ de Rennes I, until Nov 2017]
Olivier Ruas [Inria]

Interns

Matthew Atkinson [Univ de Rennes I, Jun 2017]
Louison Gitzinger [Inria, from Mar 2017 until Sep 2017]
Stewart Grant [Univ de Rennes I, from May 2017 until Aug 2017]
Hayk Saribekyan [Inria, from Jun 2017 until Aug 2017]
Jodi Spacek [Inria, from May 2017 until Aug 2017]

Administrative Assistant

Virginie Desroches [Inria]

Visiting Scientist

Arvid Bosk [KTH Royal Institute of Technology]

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 such 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 [46], [40], [41]. 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 [47], [48], [49] following a specific topology or an unstructured network [45], [50] 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 [44], [43], 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 [43], 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.) are known both for synchronous and augmented asynchronous systems (recall that in purely asynchronous systems some problems are impossible to solve). 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. Awards

- Anne-Marie Kermarrec received the Inria/Dassault Systems/Académie des science/ Innovation Award in 2017.

5. New Software and Platforms

5.1. WebGC

Web-based Gossip Communication
KEYWORDS: WebRTC - Recommendation systems - Decentralized architectures - Personalized systems - Web - Peer-to-peer - Gossip protocols - Epidemic protocols - Decentralized web
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 consists of a WebRTC-based library that supports gossip-based communication between web browsers and enables them to operate with Node-JS applications. WebGC comprises the implementation of standard gossip protocols such as Peer Sampling or Clustering, and simplifies the development of new protocols. It comprises a decentralized signaling service that makes it easier to build completely decentralized browser-based applications.

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

5.2. Asapknn (MediEgo)

KEYWORDS: Widget web - Social network - Recommendation
FUNCTIONAL DESCRIPTION: Asapknn (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: Anne Marie Kermarrec, Antoine Boutet, Arnaud Jegou, Davide Frey, Jacques Falcou, Jean-Francois Verdonck, Rachid Guerraoui and Sébastien Campion
- Partner: EPFL - Ecole Polytechnique Fédérale de Lausanne
- Contact: Sébastien Campion

5.3. YALPS

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

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

5.4. 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.

RELEASE FUNCTIONAL DESCRIPTION: Library for gossip-based applications and experiments

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

6. New Results

6.1. Theory of Distributed Systems

6.1.1. Simulation of Partial Replication in Distributed Transactional Memory

Participant: François Taïani.

Distributed Transactional Memory (DTM) is a concurrency mechanism aimed at simplifying distributed programming by allowing operations to execute atomically, mirroring the well-known transaction model of relational databases. DTM can play a fundamental role in the coordination of participants in mobile distributed applications. Most DTM solutions follow a full replication scheme, in spite of recent studies showing that partial replication approaches can present gains in scalability by reducing the amount of data stored at each node. This work [33] investigates the role of replica location in DTMs. The goal is to understand the effect of latency on the DTM's system performance in face of judicious replica distribution, taking into consideration the locations where data is more frequently accessed.

This work was performed in collaboration with with Diogo Lima and Hugo Miranda from the University of Lisbon (Portugal).

6.1.2. Distributed Universal Constructions: a Guided Tour

Participant: Michel Raynal.

The notion of a universal construction is central in computing science: the wheel has not to be reinvented for each new problem. In the context of n-process asynchronous distributed systems, a universal construction is an algorithm that is able to build any object defined by a sequential specification despite the occurrence of up to (n - 1) process crash failures. Michel Raynal presented a guided tour of such universal constructions in the bulletin of the EATCS [22]. Its spirit is not to be a catalog of the numerous constructions proposed so far, but a (as simple as possible) presentation of the basic concepts and mechanisms that constitute the basis these constructions rest on.

6.1.3. Atomic Read/Write Memory in Signature-Free Byzantine Asynchronous Message-Passing Systems

Participant: Michel Raynal.

This work introduced a signature-free distributed algorithm which builds an atomic read/write shared memory on top of a fully connected peer-to-peer 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 [42], 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 Mostéfaoui, Matoula Petrolia and Claude Jard from the University of Nantes and was published in Theory of Computing Systems [19].

6.1.4. From wait-free to arbitrary concurrent solo executions in colorless distributed computing

Participant: Michel Raynal.

In an asynchronous distributed system where any number of processes may crash, a process may have to run solo, computing its local output without receiving any information from other processes. In the basic shared memory system where the processes communicate through atomic read/write registers, at most one process may run solo.

In this work we introduced a new family of d-solo models, where d-processes may concurrently run solo, 1≤d≤n (the 1-solo model is the basic read/write model). We studied distributed colorless computations in the d-solo models, where process ids are not used, either in task specifications or during computation, and we characterized the colorless tasks that can be solved in each d-solo model. Colorless tasks include consensus, set agreement and many other previously studied tasks. This shows that colorless algorithms have limited computational power for solving tasks, only when d>1. When d=1, colorless algorithms can solve the same tasks as algorithms that may use ids. It is well-known that, while consensus is not wait-free solvable in a model where at most one process may run solo, ϵ-approximate agreement is solvable. In a d-solo model, the fundamental solvable task is (d,ϵ)-solo approximate agreement, a generalization of ϵ-approximate agreement. Indeed, (d,ϵ)-solo approximate agreement can be solved in the d-solo model, but not in the (d+1)-solo model.

This work was carried out in collaboration with Maurice Herlihy from Brown University, Sergio Rajsbaum from UNAM (Mexico), and Julien Stainer from EPFL, in the context of the LIDICo associate team. It was published in Theoretical Computer Science [18].

6.1.5. Early Decision and Stopping in Synchronous Consensus: A Predicate-Based Guided Tour

Participant: Michel Raynal.

Consensus is the most basic agreement problem encountered in fault-tolerant distributed computing: each process proposes a value and non-faulty processes must agree on the same value, which has to be one of the proposed values. While this problem is impossible to solve in asynchronous systems prone to process crash failures, it can be solved in synchronous (round-based) systems where all but one process might crash in any execution. It is well-known that $(t + 1)$ rounds are necessary and sufficient in the worst case execution scenario for the processes to decide and stop executing, where $t < n$ is a system parameter denoting the maximum number of allowed process crashes and n denotes the number of processes in the system. Early decision and stopping considers the case where $f < t$ processes actually crash, f not being known by processes. It has been shown that the number of rounds that have to be executed in the worst case is then $\min(f + 2, t + 1)$. In this work we showed that this value is an upper bound attained only in worst execution scenarios. This work resulted from a collaboration with Armando Castaneda from UNAM, Yoram Moses from Technion, and Matthieu Roy from LAAS Toulouse, in the context of the LIDICo associate team. It was published at NETYS 2017 [29].

6.1.6. Long-Lived Tasks

Participant: Michel Raynal.

The predominant notion for specifying problems to study distributed computability are tasks. Notable examples of tasks are consensus, set agreement, renaming and commit-adopt. The theory of task solvability is well-developed using topology techniques and distributed simulations. However, concurrent computing problems are usually specified by objects. Tasks and objects differ in at least two ways. While a task is a one-shot problem, an object, such as a queue or a stack, typically can be invoked multiple times by each process. Also, a task, defined in terms of sets, specifies its responses when invoked by each set of processes concurrently, while an object, defined in terms of sequences, specifies the outputs the object may produce when it is accessed sequentially.

In this work we showed how the notion of tasks can be extended to model any object. A potential benefit of this result is the use of topology, and other distributed computability techniques to study long-lived objects. This work resulted from a collaboration with Armando Castaneda and Sergio Rasjbaum from UNAM in the context of the LIDICo associate team. It was published at NETYS 2017 [35].

6.1.7. Which Broadcast Abstraction Captures k-Set Agreement?

Participant: Michel Raynal.

It is well-known that consensus (one-set agreement) and total order broadcast are equivalent in asynchronous systems prone to process crash failures. Considering wait-free systems, we addressed and answered the following question: which is the communication abstraction that "captures" k-set agreement? To this end, we introduced a new broadcast communication abstraction, called k-BO-Broadcast, which restricts the disagreement on the local deliveries of the messages that have been broadcast (1-BO-Broadcast boils down to total order broadcast). Hence, in this context, k=1 is not a special number, but only the first integer in an increasing integer sequence. This establishes a new "correspondence" between distributed agreement problems and communication abstractions, which enriches our understanding of the relations linking fundamental issues of fault-tolerant distributed computing. This work was carried out in collaboration with Damien Imbs from the University of Marseille, Achour Mostéfaoui from the University of Nantes, and Matthieu Perrin from IMDEA (Spain). It was published at DISC 2017 [39].

6.1.8. Signature-free asynchronous Byzantine systems: from multivalued to binary consensus with t< n/3, O(n2) messages, and constant time.

Participant: Michel Raynal.

We introduced 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 $t < n/3t < n/3$ (and is consequently optimal from a resilience point of view), uses $O(n^2)$ messages, has a constant time complexity, and uses neither signatures nor additional computational power (such as random numbers, failure detectors, additional scheduling assumption, or additional synchrony assumption). 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 satisfying the following property: if the set of a non-faulty process is a singleton containing value v, the set of any non-faulty process contains v. 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 $t < n/3t < n/3$ Byzantine processes, and without using cryptography techniques. Moreover, this reduction algorithm uses a single instance of the underlying binary consensus, and tolerates message re-ordering by Byzantine processes. This work, done in collaboration with Achour Mostefaoui from LS2N (Nantes), appeared in Acta Informatica [20].

6.1.9. A distributed leader election algorithm in crash-recovery and omissive system

Participant: Michel Raynal.

We introduced a new distributed leader election algorithm for crash-recovery and omission environments. Contrary to previous works, our algorithm tolerates the occurrence of crash-recoveries and message omissions to any process during some finite but unknown time, after which a majority of processes in the system remains up and does not omit messages. This work, done in collaboration with Christian Fernández-Campusano, Mikel Larrea, and Roberto Cortiñas from UPV/EHU, Spain, appeared in Information Processing Letters 2017 [16].

6.1.10. Providing Collision-Free and Conflict-Free Communication in General Synchronous Broadcast/Receive Networks

Participants: Michel Raynal, François Taïani.

This work [26] considers the problem of communication in dense and large scale wireless networks composed of resource-limited nodes. In this kind of networks, a massive amount of data is becoming increasingly available, and consequently implementing protocols achieving error-free communication channels constitutes an important challenge. Indeed, in this kind of networks, the prevention of message conflicts and message collisions is a crucial issue. In terms of graph theory, solving this issue amounts to solve the distance-2 coloring problem in an arbitrary graph. The work presents a distributed algorithm providing the processes with such a coloring. This algorithm is itself collision-free and conflict-free. It is particularly suited to wireless networks composed of nodes with communication or local memory constraints.

This work was performed in collaboration with Abdelmadjid Bouabdallah and Hicham Lakhlef from Université Technologique de Compiègne (France).

6.1.11. Randomized abortable mutual exclusion with constant amortized RMR complexity on the CC model.

Participant: George Giakkoupis.

In [30], we presented an abortable mutual exclusion algorithm for the cache-coherent (CC) model with atomic registers and CAS objects. The algorithm has constant expected amortized RMR complexity in the oblivious adversary model and is deterministically deadlock-free. This is the first abortable mutual exclusion algorithm that achieves $o(\log n/\log\log n)$ RMR complexity.

This work was done in collaboration with Philipp Woelfel (University of Calgary).

6.2. Network and Graph Algorithms

6.2.1. Tight bounds on vertex connectivity under 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 [15], we proved the analogous result for vertex connectivity, when either vertices or edges are sampled. We showed 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. If edges are sampled with probability $p = \Omega(\log(n)/k)$then the sampled subgraph has vertex connectivity $\Omega(kp)$, with high probability. Both bounds are existentially optimal.

This work was done in collaboration with Keren Censor-Hillel (Technion), Mohsen Ghaffari (MIT), Bernhard Haeupler (Carnegie Mellon University), and Fabian Kuhn (University of Freiburg).

6.2.2. Tight bounds for coalescing-branching random walks on regular graphs

Participant: George Giakkoupis.

A *coalescing-branching random walk (Cobra)* is a natural extension to the standard random walk on a graph. The process starts with one pebble at an arbitrary node. In each round of the process every pebble splits into k pebbles, which are sent to k random neighbors. At the end of the round all pebbles at the same node coalesce into a single pebble. The process is also similar to randomized rumor spreading, with each informed node pushing the rumor to k random neighbors each time it receives a copy of the rumor. Besides its mathematical interest, this process is relevant as an information dissemination primitive and a basic model for the spread of epidemics.

In [25] we studied the *cover time* of Cobra walks, which is the time until each node has seen at least one pebble. Our main result is a bound of $O(\phi^{-1}\log n)$ rounds with high probability on the cover time of a Cobra walk with $k = 2$ on any regular graph with n nodes and conductance ϕ. This bound improves upon all previous bounds in terms of graph expansion parameters. Moreover, we showed that for any connected regular graph the cover time is $O(n\log n)$ with high probability, independently of the expansion. Both bounds are asymptotically tight.

This work was done in collaboration with Petra Berenbrink (University of Hamburg), Peter Kling (University of Hamburg).

6.3. Scalable Systems

6.3.1. Agar: A Caching System for Erasure-Coded Data

Participants: Anne-Marie Kermarrec, François Taïani.

Erasure coding is an established data protection mechanism. It provides high resiliency with low storage overhead, which makes it very attractive to storage systems developers. Unfortunately, when used in a distributed setting, erasure coding hampers a storage system's performance, because it requires clients to contact several, possibly remote sites to retrieve their data. This has hindered the adoption of erasure coding in practice, limiting its use to cold, archival data. Recent research showed that it is feasible to use erasure coding for hot data as well, thus opening new perspectives for improving erasure-coded storage systems. In this work [32], we address the problem of minimizing access latency in erasure-coded storage. We propose Agar—a novel caching system tailored for erasure-coded content. Agar optimizes the contents of the cache based on live information regarding data popularity and access latency to different data storage sites. Our system adapts a dynamic programming algorithm to optimize the choice of data blocks that are cached, using an approach akin to " Knapsack " algorithms. We compare Agar to the classical Least Recently Used and Least Frequently Used cache eviction policies, while varying the amount of data cached between a data chunk and a whole replica of the object. We show that Agar can achieve 16% to 41% lower latency than systems that use classical caching policies.

This work was performed in collaboration with from Raluca Halalai and Pascal Felber from Université de Neuchâtel (Switzerland).

6.3.2. Filament: A Cohort Construction Service for Decentralized Collaborative Editing Platforms

Participants: Resmi Ariyattu Chandrasekharannair, François Taïani.

Distributed collaborative editors allow several remote users to contribute concurrently to the same document. Only a limited number of concurrent users can be supported by the currently deployed editors. A number of peer-to-peer solutions have therefore been proposed to remove this limitation and allow a large number of users to work collaboratively. These approaches however tend to assume that all users edit the same set of documents, which is unlikely to be the case if such systems should become widely used and ubiquitous. In this work [24] we discuss a novel cohort-construction approach that allow users editing the same documents to rapidly find each other. Our proposal utilises the semantic relations between peers to construct a set of self-organizing overlays to route search requests. The resulting protocol is efficient, scalable, and provides beneficial load-balancing properties over the involved peers. We evaluate our approach and compare it against a standard Chord based DHT approach. Our approach performs as well as a DHT based approach but provides better load balancing.

6.3.3. Scalable Anti-KNN: Decentralized Computation of k-Furthest-Neighbor Graphs with HyFN

Participants: Simon Bouget, David Bromberg, François Taïani.

The decentralized construction of k-Furthest-Neighbor graphs has been little studied, although such structures can play a very useful role, for instance in a number of distributed resource allocation problems. In this work [27] we define KFN graphs; we propose HyFN, a generic peer-to-peer KFN construction algorithm, and thoroughly evaluate its behavior on a number of logical networks of varying sizes. 1 Motivation k-Nearest-Neighbor (KNN) graphs have found usage in a number of domains, including machine learning, recommenders, and search. Some applications do not however require the k closest nodes, but the k most dissimilar nodes, what we term the k-Furthest-Neighbor (KFN) graph. Virtual Machines (VMs) placement —i.e. the (re-)assignment of workloads in virtualised IT environments— is a good example of where KFN can be applied. The problem consists in finding an assignment of VMs on physical machines (PMs) that minimises some cost function(s). The problem has been described as one of the most complex and important for the IT industry, with large potential savings. An important challenge is that a solution does not only consist in packing VMs onto PMs — it also requires to limit the amount of interferences between VMs hosted on the same PM. Whatever technique is used (e.g. clustering), interference aware VM placement algorithms need to identify complementary workloads — i.e. workloads that are dissimilar enough that the interferences between them are minimised. This is why the application of KFN graphs would make a lot of sense: identifying quickly complementary workloads (using KFN) to help placement algorithms would decrease the risks of interferences. The construction of KNN graphs in decentralized systems has been widely studied in the past. However, existing approaches typically assume a form of "likely transitivity" of similarity between nodes: if A is close to B, and B to C, then A is likely to be close to C. Unfortunately this property no longer holds when constructing KFN graphs. As a result, these approaches are not working anymore when applied to this new problem.

This work was performed in collaboration with Anthony Ventresque from University College Dublin (Ireland).

6.3.4. Density and Mobility-driven Evaluation of Broadcast Algorithms for MANETs

Participants: Simon Bouget, David Bromberg, François Taïani.

Broadcast is a fundamental operation in Mobile Ad-Hoc Networks (MANETs). A large variety of broadcast algorithms have been proposed. They differ in the way message forwarding between nodes is controlled, and in the level of information about the topology that this control requires. Deployment scenarios for MANETs vary widely, in particular in terms of nodes density and mobility. The choice of an algorithm depends on its expected coverage and energy cost, which are both impacted by the deployment context. In this work, we are interested in the comprehensive comparison of the costs and effectiveness of broadcast algorithms for MANETs depending on target environmental conditions. We did an experimental study of five algorithms, representative of the main design alternatives. Our study reveals that the best algorithm for a given situation, such as a high density and a stable network, is not necessarily the most appropriate for a different situation such as a sparse and mobile network. We identify the algorithms characteristics that are correlated with these differences and discuss the pros and cons of each design.

This work was done in collaboration with Etienne Rivière (University of Neuchatel), Laurent Réveillère (University of Bordeaux) and appeared in ICDCS 2017

6.3.5. An Adaptive Peer-Sampling Protocol for Building Networks of Browsers

Participant: Davide Frey.

Peer-sampling protocols constitute a fundamental mechanism for a number of large-scale distributed applications. The recent introduction of WebRTC facilitated the deployment of decentralized applications over a network of browsers. However, deploying existing peer-sampling protocols on top of WebRTC raises issues about their lack of adaptiveness to sudden bursts of popularity over a network that does not manage addressing or routing. In this contribution, we introduced SPRAY, a novel random peer-sampling protocol that dynamically, quickly, and efficiently self-adapts to the network size. We evaluated SPRAY by means of simulations and real-world experiments. This demonstrated its flexibility and highlighted its efficiency improvements at the cost of small overhead. We embedded SPRAY in a real-time decentralized editor running in browsers and ran experiments involving up to 600 communicating web browsers. The results demonstrate that SPRAY significantly reduces the network traffic according to the number of participants and saves bandwidth.

This work was carried out in collaboration with Brice Nédelec, Julian Tanke, Pascal Molli, and Achour Mostéfaoui from the University of Nantes and will appear in the World Wide Web Journal [21].

6.3.6. Designing Overlay Networks for Decentralized Clouds

Participant: Marin Bertier.

Recent increase in demand for next-to-source data processing and low-latency applications has shifted attention from the traditional centralized cloud to more distributed models such as edge computing. In order to fully leverage these models it is necessary to decentralize not only the computing resources but also their management. While a decentralized cloud has various inherent advantages, it also introduces different challenges with respect to coordination and collaboration between resources. A large-scale system with multiple administrative entities requires an overlay network which enables data and service localization based only on a partial view of the network. Numerous existing overlay networks target different properties but they are built in a generic context, without taking into account the specific requirements of a decentralized cloud. In this work [34], done in collaboration with G. Tato et C. Tedeschi from the Myriads project team, we identified some of these requirements and introduced Koala, a novel overlay network designed specifically to meet them.

7. Partnerships and Cooperations

7.1. National Initiatives

7.1.1. ANR project SocioPlug

Participants: Davide Frey, Anne-Marie Kermarrec, Pierre-Louis Roman, Francois Taiani.

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.

7.1.2. DeSceNt CominLabs

Participants: Resmi Ariyattu Chandrasekharannair, Davide Frey, Michel Raynal, Francois Taiani.

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.

7.1.3. ANR Blanc project Displexity

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

The Displexity project started in 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.

7.1.4. ANR project PAMELA

Participants: Davide Frey, George Giakkoupis, Francois Taiani.

PAMELA is a collaborative ANR project involving ASAP, Inria Lille, UMPC, Mediego and Snips. The project aims at developing machine learning theories and algorithms in order to learn local and personalized models from data distributed over networked infrastructures. This project seeks to provide first answers to modern information systems built by interconnecting many personal devices holding private user data in the search of personalized suggestions and recommendations. More precisely, we will focus on learning in a collaborative way with the help of neighbors in a network. We aim to lay the first blocks of a scientific foundation for these new types of systems, in effect moving from graphs of data to graphs of data and learned models. We argue that this shift is necessary in order to address the new constraints arising from the decentralization of information that is inherent to the emergence of big data. We will in particular focus on the question of learning under communication and privacy constraints. A significant asset of the project is the quality of its industrial partners, SNIPS and MEDIEGO, who bring in their expertise in privacy protection and distributed computing as well as use cases and datasets. They will contribute to translate this fundamental research effort into concrete outcomes by developing personalized and privacy-aware assistants able to provide contextualized recommendations on small devices and smartphones.

7.1.5. ANR project OBrowser

Participants: David Bromberg, Davide Frey, Francois Taiani.

OBrowser is a collaborative ANR project involving Inria (ASAP team), the Univ. Nantes, the Bretagne Sud. University, and Orange. The project emerges from the vision of designing and deploying distributed application on millions of machines using web-enabled technologies without relying on a cloud or a central authority. OBrowser proposes to build collaborative applications through a decentralized execution environment composed of users' browsers that autonomously manages issues such as communication, naming, heterogeneity, and scalability. The introduction of browser-to-browser communication with WebRTC's Datachannel has made these scenarios closer, but today only experts can afford to tackle the technical challenges associated with large-scale browser-based deployments such as decentralized instant-messaging (Firechat) and Infrastructure-less Mission Critical Push To Talk. O'Browser aims to solve these challenges by means of a novel programming framework.

7.1.6. ANR project DESCARTES

Participants: George Giakkoupis, Michel Raynal, Francois Taiani.

DESCARTES is a collaborative ANR project involving ASAP, Labri (U. Bordeaux), Lafia (U. Paris Diderot), Vérimag (Grenoble), LIF (Marseilles), and LINA (Nantes). Despite the practical interests of reusable frameworks for implementing specific distributed services, many of these frameworks still lack solid theoretical bases, and only provide partial solutions for a narrow range of services. In this project, we argue that this is mainly due to the lack of a generic framework that is able to unify the large body of fundamental knowledge on distributed computation that has been acquired over the last 40 years. The DESCARTES project aims at bridging this gap, by developing a systematic model of distributed computation that organizes the functionalities of a distributed computing system into reusable modular constructs assembled via well-defined mechanisms that maintain sound theoretical guarantees on the resulting system. DESCARTES arises from the strong belief that distributed computing is now mature enough to resolve the tension between the social needs for distributed computing systems, and the lack of a fundamentally sound and systematic way to realize these systems.

7.1.7. ANR-ERC Tremplin project NDFUSION

Participant: George Giakkoupis.

NDFUSION is an 18-month ANR project awarded to the PI to support his preparation for his upcoming ERC grant application. The idea of intervening in a network diffusion process to enhance or retard its spread has been studied in various contexts, e.g., to increase the spread or speed of diffusion by choosing an appropriate set of seed nodes (a standard goal in viral marketing by word-of-mouth), or achieve the opposite effect either by choosing a small set of nodes to remove (a goal in immunization against diseases), or by seeding a competing diffusion (e.g., to limit the spread of misinformation in a social network). The aim of this project is to consolidate existing work under a single, comprehensive framework, and using this framework to develop new, efficient algorithms for optimizing (maximizing or minimizing) the spread of diffusion processes. Novel aspects of the project involve issues of scalability, multiple concurrent diffusions, and the use of multistage online strategies to optimize diffusions. Results from this project are likely to be relevant to many different disciplines, from network optimization in computing to disease containment in medicine.

7.2. International Initiatives

7.2.1. Inria International Labs

- Anne-Marie Kermarrec is the scientific co-chair (with Willy Zwaenepoel) of the EPFL/Inria International Lab

7.2.2. Inria Associate Teams Not Involved in an Inria International Labs

7.2.2.1. LiDiCo

Title: Aux limites du calcul réparti

International Partner (Institution - Laboratory - Researcher):

UNAM (Mexico) - Instituto de Matematicas - Sergio Rajsbaum

Start year: 2017

See also: https://sites.google.com/site/lidicoequipeassociee/

Today distributed applications are pervasive, some very successful (e.g., Internet, P2P, social networks, cloud computing), and benefit everyone, but the design and the implementation of many of them still rely on ad-hoc techniques instead of on a solid theory. The next generation of distributed applications and services will be more and more complex and demands research efforts in establishing sound theoretical foundations to be able to master their design, their properties and their implementation. This proposal is a step in this inescapable direction.

7.3. International Research Visitors

7.3.1. Visits of International Scientists

- Peter Kling (U of Hamburg) visited ASAP (hosted by G Giakkoupis), Jan 19–25.
- Emanuele Natale (Max Planck, Saarbrücken) visited ASAP (hosted by G Giakkoupis), Apr 23–29.
- Thomas Sauerwald (U of Cambridge) visited ASAP (hosted by G Giakkoupis), Aug 21–24.
- Robert Elsässer (Salzburg U) visited ASAP (hosted by G Giakkoupis), Sep 25–29.

7.3.1.1. Internships

- Jodi Spacek from University of British Columbia, Research internship from May 2017 until Aug 2017, supervised by David Bromberg.
- Stewart Grant from University of British Columbia, Research internship from May 2017 until Aug 2017, supervised by David Bromberg.
- Hayk Saribekyan from MIT, research Internship from June 2017 to August 2017.

7.3.2. Visits to International Teams

7.3.2.1. Research Stays Abroad

- Michel Raynal was at the Hong Kong Polytechnic University from 15 September to 14 October 2017,
- David Bromberg did a visit at USP - Department of Computer Science University of São Paulo, Sao Paulo, Brazil from February 22, 2017 to March 24, 2017

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events Organisation

8.1.1.1. Member of the Organizing Committees

- George Giakkoupis was a co-organizer of the Workshop on Decentralized Machine Learning, Optimization and Privacy, Lille, France, Sep 2017.

8.1.2. Scientific Events Selection

8.1.2.1. Member of the Conference Program Committees

- François Taïani served on the TPC of the 18th ACM/IFIP/Usenix International Conference on Middleware (Middleware 2017).
- François Taïani served on the TPC of the 37th IEEE International Conference on Distributed Computing Systems (ICDCS 2017), in the Distributed Operating Systems and Middleware track.
- François Taïani served on the TPC of the 16th Workshop on Adaptive and Reflective Middleware (ARM 2017).
- François Taïani served on the TPC of the 1st Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers (SERIAL 2017)
- George Giakkoupis served as a program committee member for the 31st IEEE International Parallel and Distributed Processing Symposium (IPDPS), Orlando, Florida, USA, May 2017.
- George Giakkoupis served as a program committee member for the 44th International Colloquium on Automata, Languages, and Programming (ICALP), Warsaw, Poland, July 2017.
- George Giakkoupis served as a program committee member for the 31st International Symposium on Distributed Computing (DISC), Vienna, Austria, Oct 2017.
- Davide Frey served as a program committee member for the 11th ACM International Conference on Distributed and Event-Based Systems (DEBS 2017), Barcelona, Spain, Jun 2017.
- Davide Frey served as a program committee member for the International Symposium on Computer Architecture and High Performance Computing, SBAC PAD 2017.
- Pierre-Louis Roman served in the program committee of the 7th IEEE International Symposium on Cloud and Service Computing (SC2 2017), Kanazawa, Japan, November 2017.
- Pierre-Louis Roman served in the program committee of the 10th IEEE International Conference on Service-Oriented Computing and Applications (SOCA 2017), Kanazawa, Japan, November 2017.

8.1.3. Journal

8.1.3.1. 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.

8.1.3.2. Reviewer - Reviewing Activities

- David Bromberg was reviewer for Journal of Internet Services and Applications
- Davide Frey was a reviewer for PPNA Peer-to-Peer Networking and Applications.
- Davide Frey was a reviewer for Transactions on Cloud Computing.
- Davide Frey was a reviewer for Transactions on Sensor Networks.

8.1.4. Invited Talks

- Anne-Marie Kermarrec gave a keynote talk at DISC 2017, Vienna, Austria

- Davide Frey gave an Invited talk at the "Session d'information sur les appels a` projets 2017 Big Data" at Business France, Paris, on January 6, 2017.
- David Bromberg gave an invited talk on "Dolmen: Towards the programmatic assembly of large-scale distributed systems" as a UFG Seminar, UFG - Instituto de Informática, Goiania, Brazil, February 20, 2017
- David Bromberg gave an invited talk on "Dolmen: Towards the programmatic assembly of large-scale distributed systems" as a UFG Seminar, Department of Computer Science, University of São Paulo, Sao Paulo, Brazil, February 22, 2017
- David Bromberg gave an invited talk on "Interoperability in distributed systems: past, present and future" at DISCOTEC/DIEBS, University of Neuchatel, June 21, University of Neuchatel, 2017

8.1.5. Leadership within the Scientific Community

- François Taïani has been a member of the Steering Committee of IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems (DAIS) since 2013.
- François Taïani has been a member of the Steering Committee of the ACM/IFIP/USENIX International Conference on Middleware (Middleware) since 2014.
- Anne-Marie Kermarrecc is a member of the ARCEP Prospective Board since 2015

8.1.6. Research Administration

- François Taïani has been a member of the Scientific Orientation Committee (COS) of IRISA (UMR 6074) since 2013.
- Anne-Marie Kermarrrec was chair of the ERC Consolidator Grant Pannel in 2017
- Davide Frey is "correspondant scientifique Europe" at the DPEI for Inria Rennes.
- Davide Frey is an associate member of the COST-GTRI of Inria.
- David Bromberg is responsible of International relationships of IRISA
- David Bromberg is member of the scientific committee of the Media & Networks competitiveness cluster
- David Bromberg is scientific correspondent for ICT Digital in Rennes

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

- Engineering School: David Bromberg, tools and programming languages for the Web, 48h, 2nd year of Engineering School (M1), ESIR / Université of Rennes 1, France.
- Engineering School: David Bromberg, Distributed software architecture, 24h, 2nd year of Engineering School (M1), ESIR / Université of Rennes 1, France.
- Pôle Universitaire Français: David Bromberg, Distributed software architecture, 60h, (M2), PUF / Université of Bordeaux, Vietnam.
- Engineering School: David Bromberg, Network security, 48h, 2nd year of Engineering School (M1), ESIR / Université of Rennes 1, France.
- Engineering School: David Bromberg, Cloud for the Internet of Things, 60h, 3rd year of Engineering School (M2), ESIR / Université of Rennes 1, France.
- Engineering School: David Bromberg, Internet of Things projects, 50h, 2rd year of Engineering School (M1), ESIR / Université of Rennes 1, France.
- University of Rennes 1: David Bromberg, Software engineering for the cloud, 4h, (M1), Université of Rennes 1, France.

- Engineering School: Francois Taiani, Synchronization and Parallel Programming, 48h, 2nd year of Engineering School (M1), ESIR / Univ. Rennes I, France.
- Engineering School: Francois Taiani, Distributed Systems, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.
- Engineering School: Francois Taiani, Parallel Algorithms for Big Data, 24h, 3rd year of Engineering School (M2), ESIR / Univ. Rennes I, France.
- Engineering School: Francois Taiani, Introduction to Operating Systems, 24h, 1st year of Engineering School (L3), ESIR / Univ. Rennes I, France.
- Master: Francois Taiani, Programming Technologies for the Cloud, 28h, M2, Univ. Rennes I, France.
- Master: Davide Frey, Scalable Distributed Systems, 10 hours, M1, EIT/ICT Labs Master School, Univ. Rennes I, France.
- Master: Davide Frey, Big-Data Storage and Processing Infrastructures, 10 hours, M2-SIF, Univ. Rennes I, France.

8.2.2. Supervision

PhD: Stéphane Delbruel, Towards a Decentralized Embryomorphic Storage System [12], University of Rennes 1, 27 January 2017, Francois Taiani and Davide Frey.

PhD: Resmi Ariyattu Chandrasekharannair, Towards Decentralized Federations for Plug-based Decentralized Social Networks [11], University of Rennes 1, 5 July 2017, Francois Taiani.

PhD in progress : Pierre-Louis Roman, Epidemic Distributed Convergence for Decentralized Social Networking, Oct 2014, Francois Taiani and Davide Frey.

PhD in progress : Olivier Ruas, Dynamic Learning and Recommendations in Very Large Distributed Computing Infrastructures, Oct 2015, Anne-Marie Kermarrec and Francois Taiani.

PhD in progress: Simon Bouget, EMILIO: Emergent Middleware for Extra-Large-Scale Self Adaptation, Sep 2014, Francois Taiani and David Bromberg (since Sep 2015).

PhD in progress: Adrien Luxey, Towards New Solutions to Build Large Scale Distributed Applications in the Cloud, Oct 2016, David Bromberg.

PhD in progress: Louison Gitzinger, "Combattre les menaces à la confidentialité des données des utilisateurs sur Android." , Oct 2017, David Bromberg.

8.2.3. Juries

- François Taïani served as external examiner on the PhD defense committee of Gowri Sankar Ramachandran on 12 June 2017 at KU Leuven (Belgium)
- François Taïani as served examiner on the PhD defense committee of Nicolas Keriven on 12 October 2017 at University of Rennes 1 (France)
- François Taïani served as reviewer on the PhD defense committee of Joachim Queireix on 21 November 2017 at Université de Bordeaux (France)
- François Taïani served as chair on the PhD defense committee of Florian Grandhomme on 23 November 2017 at Université de Rennes 1 (France)
- François Taïani served as chair on the PhD defense committee of Orçun Yildiz on 8 December 2017 at École Normale Supérieure de Rennes (France)
- Anne-Marie Kermarrec served as an examiner in the HDR defense of Christian Grothoff on October 18, 2017.

8.3. Popularization

- Davide Frey published an article on the Telecom ParisTech Magazine. Issue 185.

- Pierre-Louis Roman gave an Invited talk at the Regional council of notaries of Brittany, Rennes, France, December 15, 2017. Title “Demystifying the blockchain and its utility for notaries” (talk in French).

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

[4] 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

[5] 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

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

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

[8] 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

[9] A. MOSTEFAOUI, H. MOUMEN, M. RAYNAL. *Signature-Free Asynchronous Binary Byzantine Consensus with* $t< n/3$*,* $O(n^2)$ *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

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

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] R. ARIYATTU. *Towards federated social infrastructures for plug-based decentralized social networks*, Université Rennes 1, July 2017, https://tel.archives-ouvertes.fr/tel-01622349

[12] S. DELBRUEL. *Towards an architecture for tag-based predictive placement in distributed storage systems*, Université Rennes 1, January 2017, https://tel.archives-ouvertes.fr/tel-01523568

Articles in International Peer-Reviewed Journals

[13] Z. BOUZID, M. RAYNAL, P. SUTRA. *Anonymous obstruction-free (n,k)-set agreement with n-k+1 atomic read/write registers*, in "Distributed Computing", May 2017, pp. 1 - 19 [*DOI :* 10.1007/S00446-017-0301-7], https://hal.archives-ouvertes.fr/hal-01680833

[14] A. CASTAÑEDA, C. DELPORTE-GALLET, ·. HUGUES FAUCONNIER, S. RAJSBAUM, M. RAYNAL. *Making Local Algorithms Wait-Free: the Case of Ring Coloring*, in "Theory of Computing Systems", May 2017, pp. 1-22 [*DOI :* 10.1007/S00224-017-9772-Y], https://hal.archives-ouvertes.fr/hal-01672723

[15] K. CENSOR-HILLEL, M. GHAFFARI, G. GIAKKOUPIS, B. HAEUPLER, F. KUHN. *Tight Bounds on Vertex Connectivity Under Sampling*, in "ACM Transactions on Algorithms", May 2017, vol. 13, n^{O} 2, pp. 19:1 - 19:26 [*DOI :* 10.1145/3086465], https://hal.inria.fr/hal-01635743

[16] C. FERNÁNDEZ-CAMPUSANO, M. LARREA, R. CORTIÑAS, M. RAYNAL. *A distributed leader election algorithm in crash-recovery and omissive systems*, in "Information Processing Letters", February 2017, vol. 118, pp. 100 - 104 [*DOI :* 10.1016/J.IPL.2016.10.007], https://hal.inria.fr/hal-01660694

[17] R. GUERRAOUI, A.-M. KERMARREC, T. LIN, R. PATRA. *Heterogeneous Recommendations: What You Might Like To Read After Watching Interstellar*, in "Proceedings of the VLDB Endowment (PVLDB)", June 2017, vol. 10, n^{O} 10, pp. 1070 - 1081 [*DOI :* 10.14778/3115404.3115412], https://hal.inria.fr/hal-01660746

[18] M. HERLIHY, S. RAJSBAUM, M. RAYNAL, J. STAINER. *From wait-free to arbitrary concurrent solo executions in colorless distributed computing*, in "Theoretical Computer Science", June 2017, vol. 683, pp. 1 - 21 [*DOI :* 10.1016/J.TCS.2017.04.007], https://hal.inria.fr/hal-01660566

[19] A. MOSTEFAOUI, M. PETROLIA, M. RAYNAL, C. JARD. *Atomic Read/Write Memory in Signature-Free Byzantine Asynchronous Message-Passing Systems*, in "Theory of Computing Systems", May 2017, vol. 60, n^{O} 4, pp. 677 - 694 [*DOI :* 10.1007/S00224-016-9699-8], https://hal.inria.fr/hal-01661008

[20] A. MOSTEFAOUI, M. RAYNAL. *Signature-free asynchronous Byzantine systems: from multivalued to binary consensus with t < n/3, O(n 2) messages, and constant time*, in "Acta Informatica", August 2017, vol. 54, n^{O} 5, pp. 501 - 520 [*DOI :* 10.1007/S00236-016-0269-Y], https://hal.inria.fr/hal-01660687

[21] B. NÉDELEC, J. TANKE, P. MOLLI, A. MOSTEFAOUI, D. FREY. *An Adaptive Peer-Sampling Protocol for Building Networks of Browsers*, in "World Wide Web", 2017, vol. 25, 1678 p. [*DOI :* 10.1007/S11280-017-0478-5], https://hal.inria.fr/hal-01619906

[22] M. RAYNAL. *Distributed Universal Constructions: a Guided Tour*, in "Bulletin of the EATCS", 2017, vol. 121, pp. 1-34, https://hal.inria.fr/hal-01661004

International Conferences with Proceedings

[23] F. ANDRÉ, A.-M. KERMARREC, N. LE SCOUARNEC. *Accelerated Nearest Neighbor Search with Quick ADC*, in "ICMR 2017 - ACM International Conference on Multimedia Retrieval", Bucarest, Romania, Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, June 2017, pp. 159-166, https://arxiv.org/abs/1704.07355 [*DOI :* 10.1145/3078971.3078992], https://hal.inria.fr/hal-01660730

[24] R. ARIYATTU, F. TAÏANI. *Filament: A Cohort Construction Service for Decentralized Collaborative Editing Platforms*, in "DAIS 2017 - 7th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems", Neuchâtel, Switzerland, June 2017, pp. 146-16 [*DOI :* 10.1007/978-3-319-59665-5_11], https://hal.inria.fr/hal-01617214

[25] P. BERENBRINK, G. GIAKKOUPIS, P. KLING. *Tight Bounds for Coalescing-Branching Random Walks on Regular Graphs*, in "SODA 2018 - Proceedings of the 29th ACM-SIAM Symposium on Discrete Algorithms", New Orleans, United States, January 2018, https://hal.inria.fr/hal-01635757

[26] A. BOUABDALLAH, H. LAKHLEF, M. RAYNAL, F. TAÏANI. *Providing Collision-Free and Conflict-Free Communication in General Synchronous Broadcast/Receive Networks*, in "AINA 2017 - 31st IEEE International Conference on Advanced Information Networking and Applications", Taipei, Taiwan, March 2017, pp. 399 - 406 [*DOI :* 10.1109/AINA.2017.39], https://hal.inria.fr/hal-01620356

[27] S. BOUGET, D. BROMBERG, F. TAÏANI, A. VENTRESQUE. *Scalable Anti-KNN: Decentralized Computation of k-Furthest-Neighbor Graphs with HyFN*, in "DAIS 2017 - 17th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems", Neuchâtel, Switzerland, L. Y. CHEN, H. REISER (editors), Springer, June 2017, pp. 101-114 [*DOI :* 10.1007/978-3-319-59665-5_7], https://hal.inria.fr/hal-01617211

[28] A. BOUTET, A.-M. KERMARREC, N. MITTAL, F. TAÏANI. *KIFF: un algorithme de construction de graphe KNN générique, rapide et évolutif*, in "ALGOTEL 2017 - 19èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Quiberon, France, May 2017, https://hal.archives-ouvertes.fr/hal-01518587

[29] A. CASTAÑEDA, Y. MOSES, M. RAYNAL, M. ROY. *Early Decision and Stopping in Synchronous Consensus: A Predicate-Based Guided Tour*, in "International Conference on Networked Systems (NETYS)", Marrakech, Morocco, Proceedings of NETYS 2017: the 5th International Conference on Networked Systems, May 2017, vol. 96, pp. 167 - 221 [*DOI :* 10.1007/978-3-319-59647-1_16], https://hal.archives-ouvertes.fr/hal-01559723

[30] G. GIAKKOUPIS, P. WOELFEL. *Randomized Abortable Mutual Exclusion with Constant Amortized RMR Complexity on the CC Model*, in "PODC 2017 - Proceedings of the 36th ACM Symposium on Principles of Distributed Computing", Washington, DC, United States, July 2017 [*DOI :* 10.1145/3087801.3087837], https://hal.inria.fr/hal-01635734

[31] R. GUERRAOUI, A.-M. KERMARREC, M. TAZIKI. *The Utility and Privacy Effects of a Click*, in "SIGIR 2017 - Thr 40th International ACM SIGIR Conference on Research and Development in Information Retrieval", Tokyo, Japan, Proceedings of the 40th International ACM SIGIR Conference on Research and Development

in Information Retrieval, ACM, August 2017, pp. 665-674 [*DOI :* 10.1145/3077136.3080783], https://hal.inria.fr/hal-01660733

[32] R. HALALAI, P. FELBER, A.-M. KERMARREC, F. TAÏANI. *Agar: A Caching System for Erasure-Coded Data*, in "ICDCS 2017 - 37th IEEE International Conference on Distributed Computing Systems", Atlanta, GA, United States, June 2017, pp. 1-11 [*DOI :* 10.1109/ICDCS.2017.97], https://hal.inria.fr/hal-01617146

[33] D. LIMA, H. MIRANDA, F. TAÏANI. *Simulation of Partial Replication in Distributed Transactional Memory*, in "2017 Wireless Days Conference", Porto, Portugal, March 2017, pp. 54 - 59 [*DOI :* 10.1109/WD.2017.7918115], https://hal.inria.fr/hal-01620360

[34] G. TATO, M. BERTIER, C. TEDESCHI. *Designing Overlay Networks for Decentralized Clouds*, in "Future-Cloud 2017 - International Workshop on the Future of Cloud Computing and Cloud Services", Hong-Kong, China, December 2017, https://hal.inria.fr/hal-01634409

Conferences without Proceedings

[35] A. CASTAÑEDA, M. RAYNAL, S. RAJSBAUM. *Long-Lived Tasks*, in "NETYS 2017 - 5th International Conference on NETworked sYStems", Marrakech, Morocco, Springer, May 2017, vol. 10299, pp. 439-454 [*DOI :* 10.1007/978-3-319-59647-1_32], https://hal.archives-ouvertes.fr/hal-01660646

Research Reports

[36] D. FREY, R. GUERRAOUI, A.-M. KERMARREC, M. MONOD. *Live Streaming with Gossip*, Inria Rennes Bretagne Atlantique ; RR-9039, March 2017, n^{o} RR-9039, https://hal.inria.fr/hal-01479885

[37] D. IMBS, A. MOSTEFAOUI, M. PERRIN, M. RAYNAL. *Another Look at the Implementation of Read/write Registers in Crash-prone Asynchronous Message-Passing Systems (Extended Version)*, IRISA, Inria Rennes ; LS2N-University of Nantes ; Technion - Israel Institute of Technology, February 2017, https://arxiv.org/abs/1702.08176 , https://hal.archives-ouvertes.fr/hal-01476201

[38] D. IMBS, A. MOSTEFAOUI, M. PERRIN, M. RAYNAL. *Set-Constrained Delivery Broadcast: Definition, Abstraction Power, and Computability Limits*, LIF, Université Aix-Marseille ; LINA-University of Nantes ; IMDEA Software Institute ; Institut Universitaire de France ; IRISA, Université de Rennes, June 2017, https://arxiv.org/abs/1706.05267 , https://hal.archives-ouvertes.fr/hal-01540010

Other Publications

[39] D. IMBS, A. MOSTEFAOUI, M. PERRIN, M. RAYNAL. *Which Broadcast Abstraction Captures k-Set Agreement?*, December 2017, https://arxiv.org/abs/1705.04835 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01660664

References in notes

[40] 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

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

[42] H. ATTIYA, A. BAR-NOY, D. DOLEV. *Sharing Memory Robustly in Message-passing Systems*, in "J. ACM", January 1995, vol. 42, n^o 1, pp. 124–142, http://doi.acm.org/10.1145/200836.200869

[43] 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

[44] 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

[45] 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

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

[47] 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

[48] 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

[49] 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

[50] 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