INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team ASAP

As Scalable As Possible: Foundations of large-scale dynamic systems

Rennes - Bretagne Atlantique - Futurs

THEME COM

Activity Report

2007

Table of contents

1. Team

ASAP is a bi-localized project-team at INRIA*Rennes - Bretagne Atlantique (*IRISA*) and* INRIA*Saclay - Ile de France sud. ASAP has been officially created on July 1st, 2007.*

Head of project-team
Anne-Marie Kermarrec [DR, INRIA RENNES - BRETAGNE ATLANTIQUE, HdR]

Administrative assistants
Fabienne Cuyollaa [INRIA RENNES - BRETAGNE ATLANTIQUE]
Cécile Bouton [INRIA RENNES - BRETAGNE ATLANTIQUE (since September 2007)]
Marie-Jeanne Gaffar [INRIA SACLAY - ILE DE FRANCE SUD (since September 2007)]

Research scientist INRIA
Fabrice Le Fessant [CR1, INRIA SACLAY - ILE DE FRANCE SUD]
Aline Carneiro Viana [CR2, INRIA SACLAY - ILE DE FRANCE SUD]

Research scientist University Rennes 1
Achour Mostefaoui [Associate Professor (MdC), University Rennes 1, HdR]
Michel Raynal [Professor (Pr), University Rennes 1, HdR]

Research scientist Insa de Rennes
Marin Bertier [Assistant Professor (MdC), INSA Rennes]

PhD students
Xiao Bai [INSA- UTC - Chinese Council (since September 2007)]
François Bonnet [MENRT/ENS CACHAN Grant]
Yann Busnel [ENS CACHAN Grant (until September 2007),MENRT Grant (since September 2007)]
Vincent Gramoli [MENRT Grant (until December 2007)]
Kévin Huguenin [ENS CACHAN Grant (since October 2007)]
Vincent Leroy [MENRT Grant (since September 2007)]
Erwan Le Merrer [Cifre France Telecom industrial Grant (until November 2007)]
Nicolas Le Scouarnec [Cifre Thomson industrial Grant (since November 2007)]
Étienne Rivière [MENRT Grant (until December 2007)]
Corentin Travers [MENRT Grant (until December 2007)]
Gilles Trédan [MENRT Grant]

Post-doctoral fellows
Cigdem Sengul [Post-Doc INRIA SACLAY - ILE DE FRANCE SUD (since November 2007)]
Davide Frey [Post-Doc INRIA RENNES - BRETAGNE ATLANTIQUE (since November 2007)]
Sathya Peri [Post-Doc INRIA RENNES - BRETAGNE ATLANTIQUE (since September 2007)]
Guang Tan [Post-Doc INRIA RENNES - BRETAGNE ATLANTIQUE (Since November 2007)]

Technical staff
Erwan Le Merrer [Ingénieur-Expert INRIA RENNES - BRETAGNE ATLANTIQUE (Since December 2007)]
Loup Vaillant [Ingénieur INRIA SACLAY - ILE DE FRANCE SUD (September 2007-September 2008)]

Student interns
Vincent Leroy [Master student INSA (February-July 2007)]
Girraj Meena [Internship INRIA RENNES - BRETAGNE ATLANTIQUE (May-July 2007)]
Pramod Singh [Master student, INRIA SACLAY - ILE DE FRANCE SUD (June-September 2007)]

Visiting scientist
Roberto Baldoni [University La Sapienza, Rome, Italy, June-September 2007]

2. Overall Objectives

2.1. General objectives

Recent evolutions in distributed computing significantly increased the degree of uncertainty inherent to any distributed system and led to a scale shift that traditional approaches can no longer accommodate. The key to scalability in this context lies into fully decentralized and self-organizing solutions. The objective of the ASAP project team is to provide a set of abstractions and algorithms to build serverless, large-scale, distributed applications involving a large set of volatile, geographically distant, potentially mobile and/or resource-limited computing entities.

The ASAP Project-Team is engaged in research along three main themes: *Distributed computing models and abstractions*, *Peer-to-peer distributed systems and applications* and *Data management in wireless autonomic networks*. These research activities encompass both basic research, seeking conceptual advances, and applied research, to validate the proposed concepts against real applications.

2.1.1. A challenging new setting

Distributed computing was born in the late seventies when people started taking into account the intrinsic characteristics of physically distributed systems. The field then emerged as a specialized research area distinct from networks, operating systems and parallelism. Its birth certificate is usually considered as the publication in 1978 of Lamport's most celebrated paper "*Time, clocks and the ordering of events in a distributed system*" [53] (that paper was awarded the Dijkstra Prize in 2000). Since then, several high-level journals and (mainly ACM and IEEE) conferences are devoted to distributed computing. This distributed system area has continuously been evolving, following the progresses in all the abovementioned areas such as networks, computing architecture, operating systems. We believe that the changes that occurred in the past decade involve a paradigm shift that can be much more than a "simple generalization" of previous works. Several conferences such as NSDI and IEEE P2P were created in the past 5 years, acknowledging this evolution. The NSDI conference is an attempt to reassemble the networking and system communities while the IEEE P2P conference was created to be a forum specialized in peer-to-peer systems. At the same time, the EuroSys conference has been created as an initiative of the European Chapter of the ACM SIGOPS to gather the system community in Europe.

The past decade has been dominated by a major shift in scalability requirements of distributed systems and applications mainly due to the exponential growth of network technologies (Internet, wireless technology, sensor devices, etc.). Where distributed systems used to be composed of up to a hundred of machines, they now involve thousand to millions of computing entities scattered all over the world and dealing with a huge amount of data. In addition, participating entities are highly dynamic, volatile or mobile. Conventional distributed algorithms designed in the context of local area networks do not scale to such extreme configurations. Therefore, they have to be revisited to fit into this new challenging setting. Precisely, *scalability* is one of the main focus of the ASAP project-team. Our ambitious goal is to provide the algorithmic foundations of large-scale dynamic distributed systems, ranging from abstractions to real deployment.

More specifically, distributed computing as such is characterized by how a set of distributed entities, whether they are called processes, agents, sensors, peers, processors or nodes, having only a partial knowledge of many parameters involved in the system, communicate and collaborate to solve a specific problem. While parallelism and real-time deal respectively with efficiency and on-time computing, distributed computing can be characterized by the word *uncertainty*. Uncertainty used to be created by the effect of asynchrony and failures in traditional distributed systems, it is now the result of many other factors, such as process mobility, low computing capacity, network dynamicity, scale, etc. This creates a new deal that makes distributed computing more diverse and more challenging.

2.1.2. Mastering uncertainty in distributed computing

The peer-to-peer communication paradigm emerged in the early 2000s and is now one of the prevalent models to cope with the requirements of large scale dynamic distributed systems. In order to successfully manage the increasing level of uncertainty, distributed systems should now rely on the following properties:

Fully decentralized model : A fully decentralized system does not rely on any central entity to control the system. Participating entities may act both as clients and servers. The number of potential servers thus increases linearly with the size of the system, avoiding the performance bottleneck imposed by the presence of servers in traditional distributed systems. Such systems are therefore naturally protected from failures since there is no single point of failure and many services are naturally replicated.

Self-organizing capabilities : Participating entities are by essence highly dynamic as they might be disconnected, mobile or faulty. The system should be able to handle such dynamic behaviors and get automatically reorganized to face entities arrival and departure.

Local system knowledge : Individual entities behavior is based on a restricted knowledge of the system and yet the system should converge toward global properties.

The objective of the ASAP project-team is to cope efficiently with the intrinsic uncertainty of distributed systems and provide the foundations for a new family of distributed systems for which scalability and dynamicity are first class concerns, and to provide the basis for the design and the implementation of distributed algorithms suited to this new challenging setting. More specifically, our objectives are to work on the following complementary axes:

Distributed computing models and abstractions : While many protocols have been proposed dealing with dynamic large-scale systems, there is still a lack of formal definitions with respect to the underlying computing model. In this area, our objectives are to investigate distributed computing problem solvability, and define a realistic model for dynamic systems along with the related abstractions.

Customizable overlay networks for scalability : Many peer-to-peer overlay networks, organizing nodes in a logical network on top of a physical network, have been proposed in the past five years in order to deal with large-scale and dynamic behavior. Following this trend, we intend to step away from general-purpose overlay networks that have been proposed so far and build domain-specific overlays customized for the targeted application and/or functionality. Among the core functionalities that we are targeting here are efficient search, notification and content dissemination.

2.2. Models and abstractions for large-scale distributed computing

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.

2.2.1. 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).

2.2.2. 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.

2.3. Resource management in peer-to-peer overlays

Managing resources on a large scale, be them computing resources, data, events, bandwidth, requires a fully decentralized solution. Our research in this area focuses on building the relevant overlay networks to provide core functionalities of resource management and discovery. This includes broadcast, anycast, search, notification. Overlay networks organize peers in a logical network on top of an existing networking infrastructure. The system automatically and dynamically adapts to frequent peer arrivals and departures. In practice, two main classes have been designed: structured overlay networks rely on a name structure and map object keys to overlay nodes. They provide a distributed hash table functionality (DHT). While structured peer-to-peer systems initially dominated the academic research, their exact-match interface limits their flexibility and use for various applications, notably when it comes to non-exact information retrieval. At the other end of the spectrum, unstructured overlay networks connect peers randomly (or pseudo-randomly). This class of networks is dominated by broadcast-based searching techniques, where the goal has become to enforce restrictions on broadcasting so that efficiency can be guaranteed.

In the area of overlay networks, our approach is original for the following reasons.

- First of all, we step away from the traditional approaches consisting in creating overlay networks based solely on randomization. Instead, we are focusing on creating overlays taking into account application characteristics. This translates into either connecting applications objects themselves as peers (which obviously are eventually hosted on a physical computing entity), or influencing the overlay links so that the structure of the application itself can be leveraged for a better performance. In order to purchase this goal, we strongly believe that it is not possible to rely on a generic framework applicable to all potential large-scale platforms (as the Internet, grids, or wireless autonomic networks). Instead, a large scale system is an environment where constraints are imposed by the resources (potentially limited) of the participating entities. However, this does not prevent a service designed especially for an application domain, to be applied in another context. Therefore, we tightly couple the design of distributed systems to application environments.
- Second, we strongly believe in weakly-structured peer-to-peer systems and most of our projects rely on epidemic-based unstructured overlay networks. Epidemic communication models have recently started to be explored as a general paradigm to build and maintain unstructured overlay networks. The basic principle of such epidemic protocols is that periodically, each peer exchanges information with some other peers selected from a local list of neighbors. Such protocols have shown to be extremely resilient to network dynamics [52].
- Finally, we are convinced that we can greatly benefit from the experience gathered from both existing systems and theoretical models. We spend a significant amount of energy to find, gather and analyze workloads of real systems as well as developing our own platform in the context of our peer-to-peer collaborative backup platform. Similarly, we leverage the models and abstractions defined in the first theme of ASAP to provide guarantees and analysis of the protocols we develop in this area.

2.4. Peer-to-peer computing and wireless autonomic networks

In this area, we are investigating the use of peer-to-peer algorithms in wireless autonomic networked systems. At the moment, this research area is essentially studied by the network community, although many of these issues are common to the ones encountered in distributed computing such as information propagation, resource discovery, etc.

Wireless autonomic networks and peer-to-peer networks exhibit many similarities that are worth leveraging. Scale and dynamicity are among the most striking ones. The need for scalability prevents the use of any form of centralization. The dynamic nature of such networks imposes to design self-organizing solutions to be able to support churn, disconnection, mobility *etc.*. However, wireless network specificities imply major adaptations of Internet-based peer-to-peer algorithms. More specifically, the fact that the neighbourhood of a node is entirely fixed by the physical network topology, the necessity to take into account the energy consumption, and the broadcast property of the radio communication of wireless nodes have a strong impact on the algorithm design. This recent research area allows us to leverage our peer-to-peer expertise in another application domain with specific applications.

2.5. Malicious behaviors in large scale networks

A failure model is always considered and clearly stated when designing fault-tolerant applications. The most benign faults consist of processes that execute their protocol correctly before silently stopping execution. However, processes may exhibit malicious (or arbitrary) behaviors (commonly called Byzantine processes), voluntarily or not. A Byzantine process can send spurious information, send multiple information to processes, etc. Such a behavior could be due to an external attack or even to an unscrupulous person with administrative access. More generally, Byzantine processes can also cooperate to maximize the damage caused to the system. We refer to the notion of "adversary". When defining the system failure model, it is necessary to explicit the assumed adversary. For example, can the adversary delay messages exchanged among correct processes? Can the adversary delay a correct process (by jamming the system)? Can the Byzantine processes cooperate. Is the computational power of Byzantine processes "unbounded"? In such a case, the use of cryptography is useless.

Considering malicious behaviors is therefore related to fault-tolerance but it is also in the core of the security of systems. Systems security encompasses a family of mechanisms and techniques that allow to protect the system from internal and external attacks. These mechanisms control different aspects of the system (cryptography, secured links, controlled access, etc.). Protecting a distributed system, partially under the control of an adversary is an extremely challenging task. Dealing with process crashes is far from being trivial, many problems are known to be impossible in pure asynchronous systems. Assuming Byzantine processes complicates the problem even further. This is one of the hottest topics of distributed computing today.

3. Scientific Foundations

3.1. Introduction

Research activities within the ASAP Project-Team encompass several areas in the context of large-scale dynamic systems: models and abstraction, resource management in IP-based systems, and data management in wireless autonomic networks. We provide a brief presentation of some of the scientific foundations associated with them.

3.2. Models and abstractions of large-scale dynamic systems

Finding models for distributed computations prone to asynchrony and failures has received a lot of attention. A lot of research in that 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 a given failure class is equivalent to). Those 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 [54], [47], [48]. This shows that finding dynamics distributed computing models is today a "Holy Grail" whose discovery would allow a better understanding of the essential nature of dynamics systems.

3.3. Peer-to-peer overlay networks

As mentioned before, the past decade has been dominated by a major shift in scalability requirements of distributed systems and applications mainly due to the exponential growth of the Internet. A standard distributed system today is related to thousand or even millions of computing entities scattered all over the world and dealing with a huge amount of data. In this context, 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 client 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 connect 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 [55], [56], [57] following a specific topology or an unstructured network [51], [58] where participating entities are connected in a random or pseudo random fashion. In between, lie weakly structured peer-to-peer overlays where nodes are linked depending on a proximity measure providing more flexibility than structured overlays and better performance than fully unstructured ones. Proximity-aware overlays connect participating entities so that they are connected to close neighbors according to a given proximity metric reflecting some degree of affinity (computation, interest, etc.) between peers. We extensively use this approach to provide algorithmic foundations of large-scale dynamic systems.

3.4. Epidemic protocols

Epidemic algorithms, also called gossip-based algorithms [50], [49], are consistently used in our research. In the context of distributed systems, epidemic protocols are mainly used to create overlay networks and to ensure a reliable information dissemination in a large-scale distributed system. The principle underlying the technique, in analogy with the spread of a rumor among humans via gossiping, is that participating entities continuously exchange information about the system in order to spread it gradually and reliably. Epidemic algorithms have proven efficient to build and maintain large-scale distributed systems in the context of many applications such as broadcasting [49], monitoring, resource management, search, and more generally in building unstructured peer-to-peer networks.

3.5. Malicious process behaviors

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

4. Application Domains

4.1. Panorama

Keywords: *Scientific computing*, *Wireless autonomic networks*, *cooperative applications*, *large-scale computing*, *voice on IP*.

The results of the research targeted in ASAP span over a wide range of application areas ranging from Internet-based applications, Grid computing, and wireless autonomic networked systems. Most applications are nowadays distributed and we believe that many new potential applications are yet to be discovered.

To tackle our challenging goals, we focus on a few sets of applications, which we believe are representative of large-scale distributed applications. More specifically, the constraints imposed by those applications are representative of those we deal with in ASAP.

4.2. Resource management in Internet-based applications

Internet-based applications comprise a large number of applications deployed over the Internet. Such applications however share some common characteristics. First of all, a basic assumption is that participating entities are potentially able to communicate with every other entity using IP. This has a large impact on the possible structure of an overlay network. However, the characteristics of the underlying network in terms of delay and bandwidth might have to be taken into account. This model may serve as a basis to formalize overlay connectivity in such contexts where memory or power consumptions are not an issue, but latency matters.

The actual applications that we are targeting in this area are related to resource management in large-scale distributed systems. Resource might be related to data, computing power or bandwidth. Among the numerous applications fitting in this denomination, we are especially interested in collaborative storage systems, resource discovery and allocation in Grid-like environments and large-scale content distribution and indexing. Core functionalities of such applications are search, notification and dissemination. We discuss the particular case of a peer-to-peer backup system we are currently developing in more details in the next sections.

4.3. Sensor-based applications

The advances in hardware development have made possible the miniaturization of micro-electro-mechanical systems and consequently, the development of wireless sensor networks. The combination of inexpensive, autonomous, low-power sensing, and compact devices has improved the viability of deploying large and dense wireless sensor networks able to sense the physical world. By essence, such networks require fully decentralized solutions in which the load is evenly balanced in the system, merely because participating entities have limited in power, storage and communication capabilities.

As opposed to Internet-based applications, entities, here sensors, communicate through radio links and have therefore a limited communication range. This imposes hard constraints on the structure of the resulting topology. More specifically, the overlay structure is highly dependent on the physical topology. Also, sensors, if embedded in human body for example, might be mobile. They might also fail, having some limiting physical capabilities. These properties make such systems highly dynamic.

In this context, we are targeting two main applications: data monitoring and *physical databases*. In the latter applications, as opposed to software databases virtualizing the real objects, sensors embedded on objects themselves can communicate to provide similar functionalities.

5. Software

5.1. Peerple

Keywords: *Peer-to-peer backup system*, *distributed storage systems*, *open-source software*.

Participants: Fabrice Le Fessant, Anne-Marie Kermarrec, Loup Vaillant.

Contact: Fabrice Le Fessant

Licence: GPL

Status: under development

Peerple (formerly called Palabre) is a peer-to-peer client to share personal documents with friends in a secure and reliable way, and to backup these documents on these "friends" Peerple clients. This work is done in tight collaboration with Laurent Viennot and Anh-Tuan Gai (GANG project-team INRIA Paris - Rocquencourt). We have developed the client, with the following functionalities: a web interface allows friends to connect, authentify on a client and access photo albums using an AJAX interface. A server offers a DNS service for the peer-to-peer clients, and a Mail service to notify friends about the presence of new shared files. Finally, the client is able to backup incrementally files on a local hard disk. The prototype has been released at the beginning of 2007 as an open-source project for external contributions. We are now working on distributed backups and on code modularization.

5.2. MoveNPlay

Keywords: *Portable devices*, *distributed data access*.

Participant: Fabrice Le Fessant.

Contact: Fabrice Le Fessant

Licence: Proprietary

Status: under development

MoveNPlay is a specialization of Peerple, towards access from small portable devices, in particular phones on new generation networks (wireless, Edge, 3G), and will lead to the creation of a spin-off to distribute this software (so the proprietary licensing from the beginning). MoveNPlay has been presented in September 2007 at Office 2.0 in San Francisco, and is now targeting the emerging market of iPhone users, both as a native application and as a computer application for access from the iPhone.

5.3. Backup simulator

Keywords: *Peer-to-peer backup system*, *distributed storage systems*, *simulator*.

Participants: Fabrice Le Fessant, Samuel Bernard.

Contact: Fabrice Le Fessant

Licence: not decided

Status: under development

We have developed a peer-to-peer backup simulator over PeerSim to test various placement strategies for Peerple peer-to-peer backup module. An important question for peer-to-peer backup is to evaluate the the maintenance cost with respect to bandwidth. This software enables to simulate the behavior of tens of thousands of peers backuping their data among other peers in the simulated network, with different per-client patterns of failures.

5.4. Development toolkit for gossip-based applications in peer-to-peer systems

Keywords: *Gossip*, *communication framework*, *peer-to-peer*.

Participants: Vincent Gramoli, Erwan Le Merrer, Anne-Marie Kermarrec.

Contact: Vincent Gramoli, Erwan Le Merrer

Licence: CeCILL (http://www.cecill.info/index.en.html)

Status: the current version of GossiPeer is 0.1

URL http://gossipeer.gforge.INRIA.fr

GossiPeer is a development framework for gossip-based communication protocols. It provides program designers with a toolkit for developing applications on a distributed set of machines (or nodes). GossiPeer is especially suited for large-scale deployment in dynamic settings since communication among distant nodes is based on gossip-based mechanisms. Its gossip (aka. epidemic) built-in protocols are, by essence, periodic and involve message exchanges between a constant number of neighbors. Built on this gossip communication paradigm, GossiPeer provides each node at the application level with a random set of nodes taken among all system nodes. This randomness is ensured by the implementation of algorithms that appeared recently in the literature. Additional algorithms are currently under development.

GossiPeer is developed in Java for compliance purposes, and allows deployment on world-wide distributed testbeds, e.g. PlanetLab, as well as on NFS distributed testbeds, e.g. EmuLab. At a lower level, gossip-based communication includes TCP or UDP, depending on reliability and speed requirements of the overlying application.

GossiPeer has already been used at IRISA, INRIA SACLAY - ILE DE FRANCE SUD, and Cornell University for the development and deployment of three main distributed applications: Counting, Churn Measurement, and Distributed Slicing. The repository of the GossiPeer project is handled by the INRIAGforge.

5.5. Other peer-to-peer systems

Keywords: *Peer-to-peer*, *simulation*, *unstructured overlays*.

Participants: Anne-Marie Kermarrec, Erwan Le Merrer, Etienne Rivière.

Contact: Anne-Marie Kermarrec

Licence: Not defined yet

Status: under development

Several simulators were developed in Java to evaluate the proposed peer-to-peer systems.

The SizeWalker simulator provides a generic framework to simulate unstructured peer-to-peer overlays. The simulator enables to set the way the unstructured peer-to-peer overlay is built as well as the associated counting algorithm. This simulator has been used to evaluate the SizeWalker algorithm as well as two competitors.

Second, we developed a large-scale simulator for Voronet, that can handle up to millions of nodes. This simulator is implementing both the protocol and a set of tools to examine the behavior of the system under different workloads or node behaviors.

The Sub-2-Sub workload generator software has been developed in collaboration with Marteen van Steen, Vrije Universiteit in Amsterdam. It provides a workload generator for comprehensive publish and subscribe systems evaluation and comparison. This workload generator is highly configurable, is currently used to evaluate the Sub-2-Sub system [45], and will be used to evaluate several existing peer-to-peer approaches.

The Rappel software has been developed to implement the Rappel protocol, described further in this document. It consists of a core system (approx. 10 KLOC of C++) along with either a discrete event simulation framework (approx. 4 KLOC of C++) or a asynchronous communication kernel (approx. 3 KLOC of C++ and Perl). The Rappel software can be used either as a deployed system (on top of PlanetLab or another testbed) or using simulation on a single machine (up to 20,000 nodes simulated). All peers behaviors are described using real-world traces : node availability, subscription patterns, publications. Network metrics are described using real world AS topology and delay models. The Rappel system has been successfully used on top of the Planet Lab testbed and for simulation.

The RayNet simulator has been developed to simulate up to 10,000 peers and propose a visualization module for debug purposes. The RayNet simulator has been used successfully to examine RayNet protocol behavior [24].

5.6. SeNSim simulator and visualisator

Keywords: *Sensor networks*, *dissemination*, *gossip*, *simulation*.

Participants: Marin Bertier, Gilles Trédan, Yann Busnel.

Contact: Marin Bertier

Licence: Not defined yet

Status: under development

The SeNSim simulator provides a generic environment to simulate wireless sensor networks, static or mobile. SeNSim was developed using Java and allows (1) the creation of wireless sensor networks with different design characteristics (i.e., mobility, failures, and stimulus scenarios) and (2) the analyse of different kind of protocols. Some example of protocols currently implemented in SeNSim are: dissemination, geometric structuring, coordinate system construction, and gossiping.

To achieve a correct rendering of simulations, we also developed an original visualization that is particularly suited for last-minute presentations. The java GUI represents exchanged messages for a given simulation which is useful for providing a geographic overview of the system and the simulated protocol behaviour. Available on the INRIAForge : http://sensim.gforge.INRIA.fr/

6. New Results

6.1. Panorama

Our research activities range from theoretical bounds to practical protocols and implementations for large-scale distributed dynamic systems. The target applications range from Internet-based applications to wireless autonomic networks. We focus our research on two main areas: resource management and dissemination. We believe that such services are basic building blocks of many distributed applications. We also examine these services in two networking contexts: Internet and wireless sensors. These two classes of applications, although exhibiting very different behaviors and constraints, clearly require scalable solutions.

To achieve this ambitious goal, we tackle the issues both along the theoretical and practical sides of scalable distributed computing and ASAP is organized along the following themes:

1. Models and abstractions: dealing with dynamics,
2. Resource management in large-scale dynamic systems,
3. Peer-to-peer wireless autonomic networked systems.

For each of these themes, we detail the results we obtained in 2007.

6.2. Models and abstractions: dealing with dynamics

Keywords: *Leader election*, *asynchronous message-passing systems*, *decentralized system size estimation*, *distributed shared memory systems*, *failure detector*, *failure resilience*, *persistence*, *random walk*, *set-agreement*, *synchronous system*.

6.2.1. Byzantine consensus

6.2.1.1. Reducing the cost of Byzantine consensus in synchronous systems

Participant: Achour Mostefaoui.

In a system composed of n processes where at most t can exhibit a Byzantine behavior, it is known since early eighties that t need to be smaller then a third of the total number of process to make the Byzantine consensus problem decidable. Moreover, it has been proved that the minimum number of communication steps needed is $t+1$ in the worst case. Yet, this protocol is extremely costly in terms of the size of messages and local computation (this protocol is called EIG for Exponential Information Gathering). So far, lowering this cost has led to consider smaller values of t (such as $t < n/4$) and an increased latency ($2t+2$ steps). One protocol exist with a reasonable (polynomial) cost with the initial number of steps but is extremely complex. The goal of this work is to design a simple algorithm that is as simple and resilient as the EIG protocol but with a quadratic number of steps.

6.2.1.2. Byzantine consensus with very few synchronous links

Participants: Achour Mostefaoui, Gilles Trédan.

Consensus, i.e. agreement of processes on a single value, has been a fascinating theoretical problem for decades. The aim is to understand the underlying conditions of agreement in asynchronous networks, and to provide real systems with a useful building block for designing distributed applications. Resilience is crucial for such a block to be useful. To achieve this resilience, it is important to rely on as few hypotheses as possible, for example on the crash model. Most of literature on consensus considered only the “fail-stop crash model”. As mentioned before, a more general model could be considered such as ’Byzantine fault model’.

Consensus is a hard problem in asynchronous system with a Byzantine fault model. Moreover, algorithms solving it are often extremely complex. We designed and proved a simple algorithm that allows, using authentication, the consensus to be solved in systems with no more than t faulty Byzantine process (with $t < n/3$), as soon as a process exhibits at least $2t$ synchronous links. We conjecture this is a lower bound. This work was published at the ACM PODC’07 [41] and the OPODIS’07 conferences.

6.2.2. Eventual leader service for asynchronous shared memories

Participants: Michel Raynal, Gilles Trédan.

Designing a distributed application remains a complex process. Easing this task is a challenge. One direction is to provide the programmer with useful methods for solving classical problems he may face in distributed systems. One of these classical problems is to provide all processes with the same identity of a correct process: electing a leader.

Despite a large literature on electing a leader in message passing context (processes exchange messages), few papers tackle this problem in the asynchronous shared memory context (processes read and write a common memory). Though, with the avenue of multicore processors and commodity disks, the demand on functionalities in the shared memory context is raising.

We studied this problem, exhibiting some particularly weak assumptions that allows a leader to be eventually elected. We provided two algorithms that use these assumptions to solve the problem, using only bounded variables for one, and with a lower-bounded cost on the number of write operations issued on critical registers for the other one. This work was performed in collaboration with A. Fernandez and E. Jimenez, both from the Universidad Rey Juan Carlos in Spain. It was also published at the IEEE ISORC’07 [30] symposium.

6.2.3. Super-peers identification refined: the distributed slicing problem

Participants: Vincent Gramoli, Anne-Marie Kermarrec, Michel Raynal.

In contrast to the client/server approach, peer-to-peer systems originally consider nodes to have equal capabilities and roles. File sharing applications, for example, have showed the power of heterogeneity by identifying nodes with extra capabilities and making them play different roles. For the more general purpose of taking benefit of heterogeneity in dynamic systems, we investigated the solutions to the distributed slicing problem [28]. The distributed slicing service has been proposed to allow for an automatic partitioning of peer-to-peer networks into groups (slices) that represent a controllable amount of some resource and that are also relatively homogeneous with respect to that resource. In this work, we propose two gossip-based algorithms to solve the distributed slicing problem. The first algorithm speeds up an existing algorithm sorting a set of uniform random numbers. The second algorithm statistically approximates the rank of nodes in the ordering. The scalability, efficiency and resilience to dynamics of both algorithms rely on their gossip-based models.

6.2.4. Probabilistic scalable guarantees

Participants: Michel Raynal, Vincent Gramoli, Erwan Le Merrer.

Distributed systems are now both very large and highly dynamic. This work presents probabilistic solutions to data persistence and data consistency that both achieves scalability. The first one aims at presenting a way to ensure data persistence despite dynamism, the second one defines a memory based on timed-quorum system. While the challenge of organizing peers in an overlay network has generated a lot of interest leading to a large number of solutions, maintaining critical data in such a network remains an open issue. In [34], we defined the portion of nodes and frequency one has to probe, given the churn observed in the system, in order to achieve a given probability of maintaining the persistence of some critical data. More specifically, we provided an accurate result relating the size and the frequency of the probing set along with its proof as well as an analysis of the way of leveraging such an information in a large-scale, dynamic, distributed system.

More precisely, we looked for solutions to data consistency problem applied in dynamic large-scale systems. The work presented in [32], [33] identifies the tradeoff between load-balancing and operation latency when trying to implement an atomic memory using a structured overlay. In contrast in [46], [35], we present a Timed Quorum System (TQS), a quorum system for large-scale and dynamic systems. TQS provides guarantees that two quorums, accessed at instances of time that are close together, intersect with high probability. We present an algorithm that implements TQS at its core and that provides operations that respect atomicity with high probability (using at its core a gossip-based system as described in [14]). This TQS implementation has quorums of size $O(\sqrt{nD})$ and expected access time of $O(\log\sqrt{nD})$ message delays, where n measures the size of the system and D is a required parameter to handle dynamism. This algorithm is shown to have complexity sub-linear in size and dynamism of the system, and hence to be scalable. It is also shown that for systems where operations are frequent enough, the system achieves the lower bound on quorum size for probabilistic quorums in static systems, and it is thus optimal in that sense.

6.2.5. Conditions for set agreement

Participants: François Bonnet, Michel Raynal.

The k-set agreement problem is a generalization of the consensus problem: considering a system made up of n processes where each process proposes a value, each non-faulty process has to decide a value such that a decided value is a proposed value, and no more than k different values are decided. While this problem cannot be solved in an asynchronous system prone to t process crashes when $t \geq k$, it can always be solved in a synchronous system; $\lfloor \frac{t}{k} \rfloor + 1$ is then a lower bound on the number of rounds (consecutive communication steps) for the non-faulty processes to decide.

The *condition-based* approach has been introduced in the consensus context. Its aim was to both circumvent the consensus impossibility in asynchronous systems, and allow for more efficient consensus algorithms in synchronous systems. We addresses the condition-based approach in the context of the k-set agreement problem.

A Technical Report is available about this subject, titled: “Conditions for Set Agreement with an Application to Synchronous Systems”.

6.2.6. Graph exploration with mobile robots

Participants: François Bonnet, Michel Raynal.

We consider robots that move synchronously on a graph. We want to find the maximum number of robots that can solve some given problems considering the two following restrictions: two robots are not able to stay on the same vertex, and two robots can not use the same edge during the same round. We also add some assumptions on the knowledge of robots: Do they know the map? Which part of the graph do they see at each round?

For the problem of Complete Exploration (each robot visits each vertex infinitely often), we characterize precisely the maximum number of robots that can solve this problem for any given graph. This work has been performed in cooperation with Roberto Baldoni and Alessia Milani, both from the University of Roma.

6.2.7. The notion of a timed register

Participant: Michel Raynal.

We have proposed a new type of shared object, called *timed register*, to design indulgent timing-based algorithms. A timed register generalizes the notion of an atomic register as follows: if a process invokes two consecutive operations on the same timed register which are a read followed by a write, then the write operation is executed only if it is invoked at most d time units after the read operation, where d is defined as part of the read operation. In this context, a timing-based algorithm is an algorithm whose correctness relies on the existence of a bound Δ such that any pair of consecutive constrained read and write operations issued by the same process on the same timed register are separated by at most Δ time units. An indulgent algorithm is an algorithm that always guarantees the safety properties, and ensures the liveness property as soon as the timing assumptions are satisfied. The usefulness of this new type of shared object is demonstrated by presenting simple and elegant indulgent timing-based algorithms that solve the mutual exclusion, ℓ-exclusion, adaptive renaming, test&set, and consensus problems. Interestingly, timed registers are universal objects in systems with process crashes and transient timing failures (i.e., they allow building any concurrent object with a sequential specification). This study suggests also connections with schedulers and contention managers.

This work has being performed in collaboration with Gadi Taubenfeld from the Interdisciplinary Center Herzliya, Israel and was published at the ACM SPAA'07 symposium [44].

6.2.8. Test&Set, adaptive renaming and set agreement: a guided visit to asynchronous computability

Participants: Michel Raynal, Coretin Travers.

An important issue in fault-tolerant asynchronous computing is the power of an object type with respect to another object type. This question has received a lot of attention, mainly in the context of the consensus problem, where a major advance has been the introduction of the consensus number notion, allowing the ranking of the synchronization power of base object types (atomic registers, queues, test&set objects, compare&swap objects, etc.) with respect to the consensus problem. This has given rise to the well-known Herlihy's hierarchy.

Due to its very definition, the consensus number notion is irrelevant for studying the respective power of object types that are too weak to solve consensus for an arbitrary number of processes (these objects are usually called subconsensus objects). Considering an asynchronous system made up of n processes prone to crash, this study addresses the power of such object types, namely, the k-test&set object type, the k-set agreement object type, and the adaptive M-renaming object type for $M = 2p - \lceil \frac{p}{k} \rceil$ and $M = \min(2p-1, p+k-1)$, where $p \leq n$ is the number of processes that want to acquire a new name. It investigates their respective power stating the necessary and sufficient conditions to build objects of any of these types from objects of any of the other types. More precisely, this study shows that (1) these object types define a strict hierarchy when $k \neq 1, n-1$, (2) they all are equivalent when $k = n-1$, and (3) they all are equivalent except k-set agreement that is stronger when $k = 1 \neq n-1$ (a side effect of these results is that that the consensus number of the renaming problem is 2.)

This work has being performed in collaboration with Eli Gafni from UCLA, USA, and was published at the IEEE SRDS'07 symposium [31].

6.3. Resource management in large-scale dynamic systems

Keywords: *Peer-to-peer content searching*, *RSS feeds*, *backup systems*, *gossip-based overlay construction*, *objects networks*, *publish and subscribe*, *random walk*, *structured overlay*, *system size estimation*.

6.3.1. Gossip-based overlay networks

6.3.1.1. Combining structured and unstructured peer-to-peer networks

Participants: Marin Bertier, Anne-Marie Kermarrec, Vincent Leroy.

As many different peer-to-peer overlay networks providing various functionalities have been proposed, it is likely that multiple overlays may be deployed over a set of nodes. Therefore a physical peer may hosts several instances of logical peers belonging to various overlay networks. In this work, we show that the co-existence of a structured peer-to-peer overlay and an unstructured one may be leveraged so that by building one overlay, the other overlay is automatically constructed as well.

More specifically, we show that the randomness provided by an unstructured gossip-based overlay may be used to build the routing tables of a structured peer-to-peer overlay and the other way around. Simulation results, comparing our approach with both a Pastry-like system and a gossip-based unstructured overlay, show that we significantly reduce the overhead while providing similar functionalities. The work has been published in the Proceedings of ICDCS 2007 conference [39]

Currently, our activity focus on the analysis of structured or unstructured overlay networks' requirements in order to define generic rules for deciding if two or more overlays can co-exist. This should take into account the number of neighbors, the correlation between neighbors, as well as the impact of the functionalities of each overlay . In addition, we are investigating how several gossip-based protocols can cohabit on the same physical network.

6.3.1.2. Epidemic-based small-world networks

Participants: François Bonnet, Anne-Marie Kermarrec, Michel Raynal.

In small-world networks, each peer is connected to its closest neighbors in the network topology, as well as to additional long-range contact(s), also called shortcut(s). In 2000, Kleinberg provided asymptotic bounds on the routing performance and showed that greedy routing in a n peer small-world network, performs in $O(n^{\frac{1}{3}})$ steps when the distance to shortcuts is chosen uniformly at random, and in $O(\log^2 n)$ when the distance to shortcuts is chosen according to a harmonic distribution in a d-dimensional mesh. Yet, we observe through experimental results that peer-to-peer gossip-based protocols achieving small-world topologies where shortcuts are randomly chosen, perform well in practice.

Kleinberg results are relevant for extremely large systems while systems considered in practise are usually of smaller size, typically under a million. In this paper, we explore the impact of Kleinberg results in the context of practical systems in the context of small-world networks. More precisely, based on the observation that, despite the fact that the routing complexity of gossip-based small-world overlay networks is not polylogarithmic (as proved by Kleinberg), this type of networks ultimately provide reasonable results in practice. This leads us to think that the asymptotic big $O()$ complexity alone might not always be sufficient to assess the practicality of a system where size is typically smaller that what the one theory targets. This work consists in refining the routing complexity measure for small-world networks. Simulation results confirm that random selection of shortcuts can achieve "practical" systems. Yet, given that Kleinberg proved that the distribution of shortcuts has a strong impact on the routing complexity when it comes to extremely large networks, even if the impact is smaller in practical systems, arises the question of leveraging this result to improve upon current gossip-based protocols. This work was presented at the ALPAGE Workshop in Lyon (20-21 June 2007) and was published at the Workshop Locality'07 co-located with PODC'07 [25] and at the OPODIS'07 conference [26].

We are currently working on the design of gossip-based protocols providing a good approximation of Kleinberg-like small-world topologies. To this end, we bias the peer sampling protocol so that the sampling is biases toward a Kleinberg distribution. Preliminary simulation results demonstrate the relevance of the proposed approach. We are currently investigating the impact of the Kleinberg-like sampling on the properties of the resulting graph such as the average path length, the clustering coefficient and the in-degree distribution.

6.3.1.3. Fair gossip

Participant: Anne-Marie Kermarrec.

In collaboration with Rachid Guerraoui and Maxime Monod from EPFL, Switzerland and Vivien Quéma from the SARDES project team at INRIAGrenoble - Rhône-Alpes, we are investigating the design of gossip protocols achieving fairness. Load-balancing is inherent in these protocols as the dissemination work is evenly spread among all nodes. Yet, large-scale distributed systems are usually heterogeneous with respect to application-dependent metrics and system capabilities such as bandwidth, CPU or storage. In practice, a blind load-balancing strategy might lead to a situation of unfairness that would significantly hamper the performance of the system.

In this work, we advocate the need to capture and leverage the inequalities in a large-scale system to improve the perceived quality of the gossip-based dissemination protocol. Our simple yet powerful algorithm, *Fair-Gossip*, ensures reliable information dissemination while dynamically adapting the load of a node according to this heterogeneity. *Fair-Gossip* relies on two main key mechanisms, themselves gossip-based, to conciliate fairness and reliability: *(i)* a gossip-based approximate snapshot protocol to estimate the global *wealth* of the system and enable each node to locally re-evaluate its load dynamically. *(ii)* a dynamic gossip-based compensation protocol used by each node to perform the adaptation without affecting the reliability and performance of the information dissemination. Our Planet Lab experiments convey the very fact that fairness can effectively be achieved without significantly hampering the resilience or the average latency of the dissemination.

6.3.2. Data management and querying needs

6.3.2.1. Peer-to-peer back-up

Participants: Anne-Marie Kermarrec, Fabrice Le Fessant, Samuel Bernard.

The storage capacity of computers has increased a lot in the past years: in the meantime, final users have started using this storage for important personal data, with the democratization of digital cameras, and professional data with the rise of telecomputing. Backuping all this data has become a new challenge for peer-to-peer systems, since these users are connected most of the time, often with large unused storage capacity on their disks, and unfortunately seldom take the time to properly save these important data.

Anne-Marie Kermarrec and Fabrice Le Fessant are currently designing a platform for a collaborative backup system, and this problem tackles a large set of problems: making the backup resilient to the large number of failures characterizing peer-to-peer networks, choosing where to backup the data, designing the protocols to place and retrieve the data from the network, while ensuring secrecy/privacy of the data. The prototype, currently developed by Fabrice Le Fessant within the Peerple open-source project, uses both a structured overlay, to localize stored data during restoration, and an unstructured overlay, to query for storage availability among neighbors. Contrary to most peer-to-peer backup systems, files are not stored separately on the overlay network, but gathered in volumes, encrypted using strong cryptography for privacy, and replicated using Reed-Solomon coding, to ensure availability even in the presence of high failure rates at a minimal extra storage cost.

From a language design point of view, a novel methodology has been developed to allow automatic encoding and decoding of message and file formats while keeping full backward compatibility with almost no extra programming cost [38].

We have also designed a new peer-to-peer protocol to check the availability of peers in a system. The protocol is resilient to liars, without collusion, and at a higher cost, to collusions of liars. The protocol is based on the diffusion of cryptographic keys inside the network from a very small set of trusted peers. Samuel Bernard has used the results of this protocol to evaluate the maintenance cost of a peer-to-peer backup with various placement strategies and different rates of failures. His results show that, by storing data on peers with a similar history, the maintenance cost can be negligible for stable peers.

This work is done in collaboration with Laurent Viennot and Anh-Tuan Gai from the GANG project-team, INRIA Paris - Rocquencourt.

6.3.2.2. Overlay construction and querying mechanisms

Participants: Anne-Marie Kermarrec, Etienne Rivière.

Three protocols aim at proposing new methods and algorithms for overlay construction and associated querying mechanisms: the Voronet, the RayNet, and the Rappel protocol.

The Voronet project aims at building a fully distributed overlay network for data-storage system with proven bounds on construction and routing costs. This work is done in collaboration with Loris Marchal (ENS Lyon) and Olivier Beaumont (CEPAGE project team, INRIABordeaux - Sud-Ouest). The idea of Voronet is to generalize the Kleinberg model where each peer in an overlay is connected to its neighbors on a grid as well as to a remote node. These remote nodes are selected according to an harmonic distribution of distances to original nodes, and permit the network to exhibit two keys properties of small-world systems: existence of short paths and navigability. VoroNet structure is based a locally computed Voronoï tessellation of the object space, in a multidimensional naming space. Each of node is linked to a small number of other nodes: these neighbors are those who share vertices in the Voronoï tessellation of the Euclidean space. These links are eventually forming the Delaunay complex of the set of elements. Each node knows also one other node through a "long link"; this node is not a neighbor in the Voronoï diagram but helps providing efficient polylogarithmic routing between any two nodes in the overlay, regardless of the distribution of nodes in space. The lengths of these long links follow a k-harmonic distribution in terms of skipped nodes in the overlay. VoroNet has been implemented in a simulation tool chain, and extended simulation results involving up to millions of nodes conveys its good properties in terms of routing efficiency and scalability. More, efficient search mechanisms were proposed that construct optimal spanning trees based only on local peers' information and using compass-based inversed routing principles. The approach has been published in the IPDPS international conference [23] and as an extended version in Étienne Rivière's PhD thesis [10].

The VoroNet approach is limited to the case where the needed dimensionality for affiliated search mechanisms is 2. Computing the Voronoï diagram accurately leads to exponentially increasing computing cost, and unbounded view sizes. The RayNet protocol tackles these inherent limitations. The basic principle behind RayNet is that computing the accurate Voronoï diagram is not mandatory to implement the associated search mechanisms, only neighbors really matter for the routing properties to be maintained.

RayNet approximates the Voronoï diagram in higher dimensions. Using gossip-based self-organization, each RayNet object obtains by peer-wise exchanges, a refined view of the system (which size is in O(d)), providing enough information for routing while avoiding the computation of the exact Voronoï cell. This is based on an innovative Voronoï cell size estimator relying on a Monte-Carlo algorithm. Routing efficiency is achieved by using the small-world peer sampling algorithm proposed by [26]. RayNet permits efficient native search capabilities for data sets with up to 8 dimensions. This work, done in collaboration with Olivier Beaumont (CEPAGE project team Bordeaux) has been published in the OPODIS international conference [24].

Rappel is a self-organizing dedicated overlay for dissemination of RSS feeds update. RSS Feeds are XML data associated to a website or any content provider. They include a set of entries, called update, along with their timestamp. RSS Feeds can be very popular (Google news, headline newspaper) or unpopular (as most personal blogs). A user can subscribe to a set of feeds using a feed reader client. This feed reader client periodically polls (i.e. every 30 minutes) the feed server to check for new updates. This may involve a high stress on a server if it hosts popular feeds, and this also involves lots of unneeded requests if the period of updates publication is higher than the polling period from the client applications.

Rappel proposes a new approach for RSS update dissemination. Rappel's key design choices are: (i) a peer-to-peer distributed system that is convenient both for rare and popular feeds; (ii) a design that handle network proximity as a mean to reduce the stress on the network that other peer-to-peer systems CDN may exhibit; (iii) fault tolerance and self-organization as a primary design goal and (iv) a more efficient rate of update discovery than the one that is achieved through direct polling, by rapidly disseminating updates through redundant dissemination trees.

Rappel organizes peers (subscribers and publishers) in a peer-to-peer resilient dissemination network. This network is optimized using the inherent correlation existing in human preferences graph, and using network proximity, that helps reducing the stress on the infrastructure. The latter is achieved by using network coordinates.

Rappel has been evaluated both by simulation and by deployment on the PlanetLab testbed. More precisely, simulations use real-world traces for both client behavior and network characterization: (1) subscriptions and publications are taken from the popular LiveJournal.com website; (2) node dynamics are derived from the Overnet file-sharing network observation; (3) the underlying network routing layer is an a real Internet topology; (4) delays are modeled according to PlanetLab observation. Deployment and simulations led to similar results. and the results proved that Rappel achieves its goal in term of network load balance, interest and physical proximity awareness, end-users metrics (delays, loads).

Rappel has been sent for evaluation to a first-tier Networked Systems conference. This work has been done with Jay A. Patel and Pr. Indranil Gupta, respectively PhD student and Assistant Professor at the University of Illinois at Urbana Champaign (UIUC), United States.

6.3.3. Monitoring overlay networks

6.3.3.1. Peer counting and sampling in overlay networks based on random walks

Participants: Anne-Marie Kermarrec, Erwan Le Merrer.

Counting the number of peers in a peer-to-peer systems has proven useful in the design of several peer-to-peer applications. However, it is hard to achieve when nodes are organized in an overlay network, and each node has only a limited, local knowledge of the whole system. We propose a generic technique, called the Sample&Collide method, to solve this problem. It relies on a sampling sub-routine which randomly returns chosen peers. Such a sampling sub-routine is of independent interest. It can be used for instance for neighbor selection by new nodes joining the system. We use a continuous time random walk to obtain such samples. The core of the method consists in gathering random samples until a target number of redundant samples are obtained. This method is inspired by the "birthday paradox".

We first worked on the improvement of the paper "Peer counting and sampling in overlay networks: random walks methods" which was published in 2006 in the PODC conference. We were invited to submit it to Distributed Computing journal; the new paper is named "Peer counting and sampling in overlay networks based on random walks", and published in that journal in November 2007 [17]. This work was performed in collaboration with Ayalvadi Ganesh (Microsoft Research) and Laurent Massoulié (Thomson Research).

6.3.3.2. A distributed churn measurement method

Participants: Vincent Gramoli, Anne-Marie Kermarrec, Erwan Le Merrer.

Churn, namely the rate of joins/leaves over time in a large-scale system, is a useful input parameter for various applications or protocols. We worked, as far as we know, on the first decentralized churn measurement solution for large scale arbitrary networks.

Our algorithm uses a predetermined period of monitoring. During this period, the nodes monitor the dynamic events occurring in the system, and locally compute a partial information about the churn. When this period ends, dedicated nodes aggregate the computed information to obtain a global estimate of the churn.

We are still working on the deployment if this algorithm on a real testbed, to emphasize its behavior on various network topologies.

6.4. Peer-to-peer wireless autonomic networked systems

Keywords: *Peer-to-peer overlays*, *coverage*, *gossip-based algorithms*, *power consumption*, *sensor networks*, *wireless networks*.

In this area, we are investigating the use of peer-to-peer algorithms in wireless, autonomic networked systems, and in particular, in resource-limited networks, as sensor network systems. We observe many similarities between these domains that we plan to leverage. Scale and dynamicity are among the most striking similarities between the two types of networks. However, wireless communication and resource-limit specificities imply major adaptations of Internet-based peer-to-peer algorithms. This new research area allows us to widen our application domain with specific applications, but above all to vary some major properties of our target system and therefore generalize our work on fully decentralized algorithms. In the following, we give a brief description of our current activities.

6.4.1. Data management

Participants: Marin Bertier, Yann Busnel, Anne-Marie Kermarrec.

SOLIST is a generic lightweight system architecture for large-scale wireless sensor networks (WSNs). SOLIST is composed of a finite set of overlays providing a common interface, with a type-based clustering. These overlays are based on an enhancement for WSN of a well-known peer-to-peer Distributed Hash Table (CAN: a Scalable Content-Addressable Network). Based on its lightweight structure, SOLIST provides an efficient implementation of a set of communication primitive (anycast, k-cast and broadcast) with respect to energy saving and reliability.

6.4.2. Software updating with gossip-based algorithms

Participants: Marin Bertier, Yann Busnel, Anne-Marie Kermarrec.

We also proposed a gossip-based algorithm for software updates in Sensor networks. This work has been done in collaboration with Éric Fleury, ARES project-team, INRIAGrenoble - Rhône Alpes. This algorithm, largely inspired from the epidemic paradigm, allows updating persistent data in large-scale WSNs with a good trade-off between data propagation speed and load balancing.

6.4.3. Information dissemination in sensor networks

Participants: Aline Carneiro Viana, Cigdem Sengul, Marin Bertier, Anne-Marie Kermarrec.

We conducted a preliminary study of information dissemination in sensor networks in the context of monitoring children activities to detect obesity pathologies. This application brings us a challenging setting with respect to dynamics as sensors are embedded on human beings and therefore mobile.

More specifically, the challenge here is the propagation of collected data from mobile sensor nodes to the sinks (monitoring station) in charge of processing it. This is a major issue in sensor-based mobile applications. In this context, the flooding of messages to the entire network should be avoided. Moreover, considering the mobility of base stations and sensor nodes, it is important to ensure that collected data will reach base stations in a reliable and robust manner. The question we intend to deal is "how to make sensors and sinks to communicate in a reliable way, without a prior absolute knowledge about their locations". We are investigating methods for energy-efficient route discovery and for the reliable relaying of data from the sensor to the base stations. While many approaches in this area rely on precise location information such as GPS, we want to avoid such techniques for cost, size but also performance reasons. We are investigating the use of epidemic algorithms to trace base station locations and efficiently transport data to them without flooding the network. The research work performed here is inserted in the context of the RNRT SVP project.

6.4.4. Self-management

6.4.4.1. Trajectory tracking of anonymous objects on graphs

Participants: Marin Bertier, Yann Busnel, Anne-Marie Kermarrec.

The aim of this problem is to track the trajectory of anonymous objects in a defined area. After enhancing and refining the system model, we propose various algorithms (centralized and distributed) to solve this problem in realistic contexts, such as tracking boats in Venice's channels.

6.4.4.2. *Virtual coordinates for autonomous networked system*

Participants: Gilles Trédan, Aline Carneiro Viana, Michel Raynal, Anne-Marie Kermarrec, Achour Mostefaoui.

The motivation behind this research work comes from the lack in the literature, of an autonomous system able (1) to permanently evolve and self-organize under dynamic changing conditions (due either to the environment or technological issues), and (2) to provide various networking functionalities over the same underlying support system. We argue that networks must not only scale in size but also in functionality. In particular, we observe that the related proposals are designed focusing on various different segment of network functionality: network slicing, virtual coordinates, data aggregation, load balancing, etc.

In general, due to the way they are designed, adding a not previously envisaged new feature requires the whole network reconfiguration and/or the provision of new node capabilities. These requirements can be, however, invalidated or deferred if the network is deployed in an area of difficult access. Our answer to those demands is an autonomous system able not only to be adaptable to environment changing conditions, but also, that provides variety to network functionalities. Thus, we are designing an autonomous and lightweight self-organizing networked system that, by imposing a bounded overhead to wireless devices, constructs a base network structure for supporting network functionalities, commonly required in WSNs. We are only exploiting local connectivity information and per-neighbor communication. Some local information that a node can infer from the wireless radio communication are: underlying connectivity, messages overhearing, and neighborhood variations. The proposed base network structure results in a virtual coordinate system allowing the network to structure itself.

6.4.4.3. *Distributing load in wireless resource-limited networks*

Participants: Vincent Gramoli, Aline Carneiro Viana, Anne-Marie Kermarrec.

In this area, we narrow our focus to a typical environment consisting in a large number of sensor nodes deployed to collect data or events in a specified geographic area and in a mobile sink moving over the region to collect monitored data. The main challenge in such a context is to safely store collected data such that they can be retrieved later on, despite the dynamism of system participants. In this context, in collaboration with Rachid Guerraoui (EPFL, Lausanne), we have been investigating a new approach that allows existent distributed abstractions to be adapted for resource-limited and dynamic sensor-based environments. Our main goals are to extend network lifetime and to evenly spread the load over the network by aggregating collected data in some selected storage nodes as well as to improve data availability by replicating aggregated data onto selected storage nodes. Following these goals, we have focused our approach on the use of quorum systems, a shared-memory concept widely adopted by the distributed system community and that constitute an important abstraction for achieving fault-tolerance, consistency, availability, and load balancing. We seek thus an optimal partition of the monitored region into quorums according to some cost measures. These measures will be the communication cost for aggregation, the load cost for replication and the network lifetime extension. We intend then to define a quorum system that can be used as a building block for gathering monitored information in a wireless sensor networks. This work can also be extended for dealing with the presence of malicious node in the network. The research work performed here is conducted in the context of the ARC Malisse project.

6.4.4.4. *Energy-efficient route discovery in sensor networks*

Participant: Aline Carneiro Viana.

The vast literature on the wireless sensor research community contains many valuable proposals for managing energy consumption, the most important factor that determines sensor lifetime. The goal of this work is to extend the network lifetime. We aim at determining good energy-efficient routes in the network by using the energy level of nodes as a criterion to select good links in the route. The estimation of node energy level in the network is not a trivial task, since in a wireless radio communication the remaining energy of a node can be affected by many factors: its data transmission, a data reception, and the interference caused by a closer data transmission. Interesting researches have been facing this requirement by focusing on the extension of the entire network lifetime: either by switching between node states (active, sleep), or by using energy efficient routing.

In collaboration with Khaldoun Al Agha and Joseph Rahme both from the LRI/Universite Paris-Sud, we are working on the definition of some energy cost models that will allow us to correctly compute the node energy in order to better estimate energy-efficient routes. This work was published in the IFIP IHN conference, the 1st Home Networking Conference [43].

6.4.4.5. Target coverage in wireless sensor networks

Participant: Aline Carneiro Viana.

This recent collaboration with Marcelo Dias de Amorim from CNRS/LIP6 focus on applications of wireless sensor networks that require periodic readings. This means that these readings should be performed following some predefined parameter $f_{\min}$ that denotes the minimum frequency at which the whole target area must be sensed. An interesting solution for this problem is to use mobile sensors that move around in order to cover multiple target regions. Previous related works addressed *why* mobility is useful in WSNs. They have shown that mobility increases coverage and reduce energy consumption for relaying traffic in wireless sensor networks.

In contrast, we also consider mobile sensor network, but focus on *how* sensors should move, in order to guarantee the coverage of all targets in the network in a timely and efficient way. This is a difficult problem, as many constraints exist in order to deal with the latency and coverage issues. The failure to visit some target points result in data loss, while the infrequent visit of points result in long delivery delays.

The solution we are working on allows to determine good/correct sensors' trajectories that (1) guarantees the coverage of the entire target area, (2) limits the number of required mobile sensors in the monitored region, and (3) bounds the delivery delay of readings.

6.4.4.6. Building secured links is sensor networks

Participants: Marin Bertier, Achour Mostefaoui, Gilles Trédan.

This work deals with malicious behaviors in the context of sensor networks. Such a behavior can be due to an adversary that has some sensors under control or more generally to a problem of the sensor itself. Effectively, as sensors are small devices that are industrially built, many of them may be defective. Moreover, it is known that when a sensor is running out of energy, it can enter a state where it abnormally behaves. Malicious behaviors in sensor networks less hard to handle as the power of the adversary is lower. Indeed a sensor has a limited energy. The more it is active the less it will survive and thus even its computation power is bounded. In the case of a sensor network with static sensors, we try to build secured links between sensors. The objective is to avoid the case of an adversary that collects the whole information exchanged among the sensors.

6.4.5. Consensus in MANET

Participant: François Bonnet.

In a loss- and disconnection-prone network, such as a mobile ad-hoc network (MANET), complete coverage requires that a message be transmitted until all operative nodes acknowledge reception. This is prohibitively expensive when small, mobile wireless devices collaborate by forming a MANET. This paper presents a solution to the consensus problem, essential to support user collaboration, using a new class of broadcasts that intentionally sacrifice full coverage for savings in memory and bandwidth. However, when such broadcasts are used, consensus amongst n devices cannot be guaranteed if at most f of them can crash (unnoticeably) and if $n \leq 3f$. We present a protocol for $n > 3f$ and evaluate its performance through simulations.

This work was published at the Workshop on Dependable Application Support for Self-Organizing Networks co-located with DSN'07 [22] and has been performed in cooperation with Khaled Alekeish and Paul Ezhilchelvan, both from the University of Newcastle.

7. Contracts and Grants with Industry

7.1. France Telecom

Participant: Anne-Marie Kermarrec.

Since October 2004, we have a collaboration with France Télécom R&D, Lannion on applying peer-to-peer techniques to telecom operator frameworks. More specifically, in this area, we are working on timely dissemination of voice over IP and a reliable and distributed telecom infrastructure. In this context, Anne-Marie Kermarrec acts as the PhD advisor of Erwan le Merrer.

7.2. Advestigo

Participants: Marin Bertier, Anne-Marie Kermarrec, Fabrice Le Fessant.

We have a consulting contract with Advestigo, a small company working in the content protection area. In this context Marin Bertier, Anne-Marie Kermarrec and Fabrice Le Fessant are providing expertise in the area of monitoring peer-to-peer systems.

8. Other Grants and Activities

8.1. National grants

8.1.1. ACI Masse de Données Alpage

Participants: Marin Bertier, Anne-Marie Kermarrec, Fabrice Le Fessant, Étienne Rivière.

ALPAGE is an ANR "Masse de Données" project started in January 2006 focusing on algorithms for large-scale platforms. The project gathers several teams with complementary expertise ranging from algorithms design and scheduling techniques, to macro-communications primitives and routing protocols and to peer-to-peer architectures and distributed systems. In this project, we aim at designing algorithms for large-scale dynamic platforms and will concentrate our efforts on the following complementary areas:

- Large-scale distributed platform modeling
- Overlay network topologies
- Scheduling for regular parallel applications
- Scheduling for file sharing applications

The partners includes the INRIABordeaux - Sud-Ouest, (contact: Olivier Beaumont), INRIALyon Rhône-Alpes (contact: Yves Robert), LRI (Contact: Pierre Fraigniaud) and INRIA RENNES - BRETAGNE ATLANTIQUE (contact: Anne-Marie Kermarrec). In this context, the ASAP project-team is mostly involved in the overlay network topologies theme and in this context we are actively collaborating with Olivier Beaumont.

8.1.2. RNRT project SVP

Participants: Marin Bertier, Yann Busnel, Anne-Marie Kermarrec, Aline Carneiro Viana.

The SVP project addresses the understanding, the conception, and the implementation of an integrated ambient architecture that would ease the optimization in the deployment of surveillance and prevention services in different types of dynamic networks. The main objective is to develop an environment which is able to accommodate a high number of dynamic entities completely dedicated to a specific service. The partners of the project come from various research communities: network, distributed system, sensor architecture and metabolical and mechanical motion control (CEA, ANACT, APHYCARE, INRIA, UPMC/LIP6, LPBEM, Thalès). Our work on sensor networks for health monitoring applications takes place in this context.

8.1.3. RNRT project SensLAb

Participants: Marin Bertier, Anne-Marie Kermarrec.

Recently accepted by the ANR, this project gathers academic and industrial partners. The purpose of this project is to deploy a very large-scale open wireless sensor network platform to be used as an efficient scientific tool for designing, tuning, and experimenting real sensor-based applications. Consequently, a SensLAB platform composed of 1024 nodes will be deployed among 4 sites. This infrastructure will represent the unique scientific tool for the research on wireless sensor networks.

8.1.4. ARC INRIA Recall

Participants: Anne-Marie Kermarrec, Achour Mostefaoui, Michel Raynal.

Anne-Marie Kermarrec, Achour Mostefaoui and Michel Raynal are involved in the ARC *Recall* on optimistic replication for collaborative editing in peer-to-peer networks. The INRIAproject-teams CASSIS and REGAL as well as the LIRMM and EPFL (Rachid Guerraoui) are involved in this project as well.

8.1.5. ARC INRIA Malisse

Participants: Anne-Marie Kermarrec, Marin Bertier, Aline Carneiro Viana, Achour Mostefaoui, Michel Raynal.

This is a collaboration project between ASAP and ARES INRIAGrenoble - Rhône-Alpes project team, and EPFL Lausanne (Rachid Guerraoui) school. The goal of this project is to explore a new generation of sensor networks in which tiny devices permanently self-organize. We focus in this project on the impact of malfunctions and misbehavior of individual devices. In fact, we believe that to be applicable to a wide range of applications, WSNs should be able to support reliably a number of key functionalities, even in the presence of malicious sensors.

8.1.6. RTRA Digiteo

Participant: Aline Carneiro Viana.

DigiteoLabs is a recently created virtual lab that has as goal to gather and promote collaborations between the following research centers: INRIA, University of Paris-Sud, Supelec, Ecole Polytechnique, and CEA. A call for regional collaborating projects was recently opened. In this context, the ASAP project entitled "Quality of Service in wireless sensor network" was selected, which will allow us to finance a 1-year Post-Doc fellowship. The project targets the resource and data management in wireless sensor networks. Khaldoun Al Agha and Steven Martin from the LRI/University of Paris-Sud are also part of the project.

8.1.7. Project RIAM-Solipsis

Participants: Davide Frey, Anne-Marie Kermarrec, Fabrice Le Fessant, Étienne Rivière.

This grant is supported by the ANR program RIAM and aims at designing a virtual world system such as *Second Life* in a fully distributed way. This project involves several partners: the ADEPT project team of INRIA RENNES - BRETAGNE ATLANTIQUE, France Télécom R&D, the LARES lab at University of Rennes 2, and the companies Archivideo and Artefacto.

This project deals with social usages, 3D-modelisation of objects in the virtual world, and the peer-to-peer infrastructure. The role of ASAP in this project is to provide the underlying peer-to-peer infrastructure and we specifically leverage here the work done in the Voronet and Raynet project as well as our work on epidemic protocols.

8.1.8. Project Pôle de Competitivité Images & Réseaux - P2Pim@ges

Participants: Anne-Marie Kermarrec, Erwan Le Merrer.

The P2Pim@ges project deals with secure multimedia file distribution in peer-to-peer environments. This grant is supported by the *Pôle de Competitivité Images & Réseaux* way and involves the following partners: Thomson R&D, Thomson Broadcast & Multimedia, Mitsubishi Electric ITE/TCL, Devoteam, France Télécom, ENST Bretagne, Marsoin, IPdiva, TMG and eOdus.

In this context, the role of the ASAP project-team is to provide the peer-to-peer infrastructure to distribute large multimedia files in an efficient way. More specifically, we are investigating the use of epidemic protocols to achieve efficient and relevant clustering to enable such functionalities.

8.2. International grants

8.2.1. ReSIST European project

Participants: Marin Bertier, Achour Mostefaoui, Michel Raynal, Corentin Travers.

ReSIST is an NoE (Network of Excellence) that addresses the strategic objective "Towards a global dependability and security framework" of the European Union's FP6 Work Programme for IST (Information Society Technologies), and responds to the stated "need for resilience, self-healing, dynamic content and volatile environments". The contract supporting the ReSIST activities extends on 3 years, starting on January 1st 2006.

ReSIST integrates leading researchers active in the multidisciplinary domains of Dependability, Security, and Human Factors, in order that Europe will have a well-focused coherent set of research activities aimed at ensuring that future "ubiquitous computing systems", the immense systems of ever-evolving networks of computers and mobile devices which are needed to support and provide Ambient Intelligence (AmI), have the necessary resilience and survivability, despite any residual development and physical faults, interaction mistakes, or malicious attacks and disruptions.

ReSIST's partners are: Budapest UTE (HG), City U. (UK), TU Darmstadt (DE), Deep Blue Srl (IT), France Télécom R&D (FR), IBM Research GmbH (CH), Institut Eurecom (FR), IRISA (FR), IRIT (FR), LAAS-CNRS (FR), Lisbon U. (PT), Newcastle upon Tyne U. (UK), Pisa U. (IT), Qinetiq (UK), Roma U. La Sapienza (IT), Ulm U. (DE), Southampton U. (UK), Vytautas Magnus U. (LT).

The current state-of-knowledge and state-of-the-art reasonably enables the construction and operation of critical systems, be they safety-critical (e.g., avionics, nuclear control) or availability-critical (e.g., back-end servers for transaction processing). The situation drastically worsens when considering large, networked, evolving, systems either fixed or mobile, with demanding requirements driven by their domain of application. There is statistical evidence that these emerging systems suffer from a significant drop in dependability and security in comparison with the former systems. There is thus a dependability and security gap opening in front of us. Filling the gap clearly needs dependability and security technologies to scale up, in order to counteract the two main drivers of the creation and widening of the gap: complexity and cost pressure.

8.2.2. Epi-Net Associated Team with Vrije Universiteit, Amsterdam, NL

Participants: Marin Bertier, François Bonnet, Anne-Marie Kermarrec, Etienne Rivière.

Epi-Net is an associated team from January 1st, 2006. Epi-Net addresses several applications using epidemic-based unstructured networks. Gossip-based communication models have recently started to be explored as a general paradigm to build and maintain unstructured overlay networks. More specifically, they have shown to provide a scalable way of implementing and maintaining highly dynamic unstructured overlays in which nodes can frequently join and leave. Many variants of such protocols exist and they mainly differ in deciding which neighbor to communicate with, deciding on exactly which neighbors to exchange information on, and, in the end, deciding on which peers to keep in the list to prevent it from growing unboundedly.

Epi-Net in this context to acknowledge the fact that gossip-based protocols are a powerful tool that can be turned into generic building blocks to build large-scale systems. Epi-Net started in 2006, following a one-year Van-Gogh PAI (2205) grant. 2007 has been a third year of active collaboration between the two groups and has actually turned into a consolidation year where many publications were generated involving members of the two groups.

Following the major event for Epi-Nets that we organized in Leiden in December 2006, the highlight of 2007 is the special issue of the ACM Operating System review: Volume 41, Number 5 October 2007, Special Topic Gossip-Based Computer Networking, guest editors Anne-Marie Kermarrec and Maarten van Steen

Since the workshop took place in December 2006, after the activity report of 2006 was written, we provide here a description of the actual workshop as well as the outcomes. The workshop was organized in December 2006 at the Lorentz Center in Leiden, The Netherlands, with a strong support of the Lorentz Center of Leiden, the European Community and INRIAthrough Epi-Net. The main goal was to get an overview of the state-of-the-art in gossiping networks, and to identify the important topics that we need to address in the near future. The format of the workshop was somewhat unusual in the sense that there were relatively few presentations, but many discussion slots. Roughly, each day started with a plenary presentation in which people were asked to be somewhat provocative. The total group of approximately 50 participants was divided into smaller groups, each concentrating on a specific theme, and which were asked to prepare a position paper. The participants were a mix of international senior researchers having worked for a long time in the field (such as Ken Birman and Robbert van Renesse from Cornell University, Roy Friedman from Technion Haifa, Israel, Ozalp Babaoglu from University of Bologna, Lorenzo Alvisi from Austin University, Rachid Guerraoui from EPFL to cite the most well-known), young researchers (junior researchers and PhD candidates) and the members of the two groups involved in Epi-Net. This led to discussions during the afternoon sessions, with plenary feedback at the end of the day. The results of those discussions have been laid down in the position papers published in the aforementioned special issue of ACM Operating System review, published in October 2007.

This year, Étienne Rivière, François Bonnet and Erwan Le Merrer visited Vrije Universiteit for a few weeks. Maarten van Steen visited ASAP in November 2007.

8.2.3. Collaboration with University of Illinois, Urbana-Champaign

Participants: Anne-Marie Kermarrec, Etienne Rivière.

In 2005 we started a 2-year collaboration with Indranil Gupta's team from the University of Illinois at Urbana Champaign (UIUC). Indranil Gupta visited INRIA in June 2006 and he initiated, in collaboration with Anne-Marie Kermarrec and Étienne Rivière, the Rappel project. Étienne Rivière visited UIUC in September for two weeks to carry on the collaboration.

8.2.4. PAI Hong Kong

Participants: Michel Raynal, Marin Bertier, Corentin Travers.

The aim of this research project is to explore the world of agreement problems in the context of MANET (Mobile Adhoc NETwork) and in the context of mobile agents. During these last two years, three visits took place. The agreement problems have been extensively studied in classical distributed systems, namely, systems where the processes communicate through a shared memory or a fixed network. Here we want to solve in the context of MANET (Mobile Adhoc NETwork) and in the context of mobile agents. What makes the problem difficult to solve is the combination of failures and asynchrony that prevents the entities to know which of them can actually actively participate in the consensus algorithm. These new contexts add new difficulties.

8.3. Visits (2006-2007)

Ken Birman Cornell University, January 2007.

Hong Va Leong Hong-Kong University, 12-14 February 2007.

Pascal Felber University of Neuchatel, Switzerland, March 2007

Leslie Lamport Microsoft Research, 18-19 June, 2007.

Roberto Baldoni University La Sapienza, Rome, Italy, June-September 2007.

Jiannong Cao Hong-Kong University, 19-22 November 2007.

9. Dissemination

9.1. Community animation

9.1.1. Leaderships and community service

A.-M. Kermarrec is a member of a CNRS group of experts on networking (*Comité d'experts réseaux*), a member of the steering committee of RESCOM (*pôle du GDR ASR du CNRS* gathering the French community interested in networking), and a member of the GDR Grid, peer-to-peer and parallelism.

9.1.2. Editorial boards, steering and program committees

A.-M. Kermarrec and M. Raynal organized a workhop on the "Future Trend of Distributed Computing" November 21st, 2007 in Rennes with the following speakers: Prof. Haggit Atya (Technion, Israel), Prof. Roberto Baldoni (University of Rome, Italy), Prof. Pascal Felber (University of Neuchatel, Switzerland), Prof. Roy Friedman (Technion, Israel), Dr. Pierre Fraigniaud (LIAFA, Paris), Prof. Rachid Guerraoui (EPFL, Switzerland), Prof Nir Shavit (University of Tel-Aviv, Israel), Prof Alexander Shvartzman (University of Connecticut, USA) and Prof Maarten van Steen (Vrije Universiteit, The Netherlands).

A.-M. Kermarrec was the General Chair (PC and Organization Chair) of Euro-Par 2007, held in Rennes in August 2007 (around 300 attendees, http://www.europar.org/).

She serves in the *Steering Committee* of the *Euro-Par* annual conference series on parallel computing

She was the Guest Editor with Maarten van Steen of the ACM SIGOPS Operating System Review, Special Issue on Gossip-based networking, October 2007.

She served in the program committees for the following conferences:

Infocom'07: *IEEE Conference on Computer Communications and Networking*, Anchorage, Alaska, USA, May 2007.

NSDI'07: *USENIX Symposium on Networked Systems Design & Implementation*, Cambridge, MA, Boston, April 2007.

ICDCS'07: *International Conference on Distributed Computing Systems* , Data Management track, Toronto, Canada, June 2007.

OPODIS'07: , Guadeloupe, France, December 2007.

P2P'07: *IEEE Conference on Peer-to-Peer systems*, Gallway, Ireland, September 2007.

SASO'07: *First IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, Boston, USA, July, 2007

HotDep'07 *the Third Workshop on Hot Topics in System Dependability* In conjunction with DSN'2007, Edinburgh, UK (June 26-28th)

CEC'07: *Special session on Evolutionnary Computing for Decentralized systems*, Singapore, September 2007.

InterPerf'07: , Nantes, France, October 2007.

IWSOS'07: *New Trends in Network Architectures and Services: International Workshop on Self-Organizing Systems*, The Lake District, UK, September, 2007

ICDE'08: *IEEE 24th International Conference on Data Engineering (ICDE 200))*, Distributed, Parallel, and Peer-to-Peer Databases Track, to be held at Cancoun Mexico, 2008.

IPDPS'08: *IEEE International Conference on Parallel and Distributed Systems*, Miami, Florida, September 2008.

EuroSys'08: to be held in Glasgow, Scottland, April 2008.

NSDI'08: *USENIX Symposium on Networked Systems Design & Implementation*, PC-Lite, San Francisco, CA, April 2008.

SASO'08: *IEEE International Conference on Self-Adaptive and Self-Organizing Systems*, Venice, Italy, October 2008.

ICDCS'08: *International Conference on Distributed Computing Systems* , Cyber-Infrastructure for Distributed Computing Track, Beijing, China, June 2008.

DEBS'08: *2nd International Conference on Distributed Event-Based Systems*, Rome, Italy, July 2008.

Aline C. Viana served in the Shadow Program Committees for CoNEXT 2007 (New York, December 2007) and in the Program Committees for IMAGINE 2007 (Lemesos, Cyprus, September 2007), the First International workshop on Mobility, Algorithms and Graph theory In dynamic Networks, associated to DISC 2007.

Fabrice Le Fessant served in the program committees for the following conferences:

Euro-Par 2007: *Peer-to-Peer and Web Computing* Topic of Euro-Par 2007, Rennes, France, August 2007.

ICFP 2007: *12th ACM SIGPLAN International Conference on Functional Programming*, Freiburg, Germany, October 2007.

Achour Mostefaoui served in the Program Committees of the following conferences:

Euro-Par 2007: *Distributed Systems and algorithms* Topic of Euro-Par 2006, to be held in Rennes, France, August 2007.

DISC'07: 21st Annual Conference on Distributed Computing, Lemesos, Cyprus, September 2007.

FOFDC'07: ARES 2007 Workshop on Foundations of Fault-tolerant Distributed Computing, Vienna, Austria April 2007.

ISPS'07: 8th International Symposium on Programming and Systems, Santa Clara, CA, June 2007.

Michel Raynal is a member of the editorial board of the journal IEEE TPDS.

He is a member of the Steering Committees of the following conferences: ACM PODC, SIROCCO and ICDCN.

He served in the program committees of the following conferences:

AINA'07: *IEEE conference on Advances Information Networking and Applications*, Niagara Falls, Canada, May 2007.

ICA3p2'07: *IEEE Conference on Algorithms and Architectures for Parallel Processing*. IEEE Computer Society Press.

ICDCS'07: *27th IEEE Int. conf. on Distributed Computing Systems*, Toronto, Canada, June 2007.

DSN'07: *37th Int'l IEEE/IFIP Conf. on Dependable Systems and Networks*, Edinburgh, Scotland, June 2007.

DASSON'07: *Workshop on Dependable Application Support in Self-Organising Networks*, in coordination with DSN'07, Edinburgh, United Kingdom, June 2007.

SRDS'07: *26th IEEE Int'l Symposium on Reliable Distributed Systems*, Beijing, China, October 2007. IEEE Computer Society Press.

SIROCCO'07: *14th Int'l Colloquium on Structural Information and Communication Complexity*. Springer-Verlag LNCS, Castiglioncello (LI), Italy, June 2007.

PaCT'07: *9th International Parallel Computing Technologies*, Romania, September 2007. Springer-Verlag, LNCS, Brasov.

Marin Bertier served in the program committees for the following conferences:

MSN'07: *The 3rd International Conference on Mobile Ad-hoc and Sensor Networks*, Beijing, China, December 2007.

Algotel 2007: *The International Conference on Dependable Systems and Networks*, Ile d' Oléron, France, Mai 2007.

Autonomics 2007: *First International Conference on Autonomic Computing and Communication Systems* to be held in Rome, Italy, October 2007.

He also serves as the Web master and Publication Chair in the Euro-Par 2007 Organisation Committee.

9.1.3. Evaluation committees, consulting

A.-M. Kermarrec served as a reviewer of the

- EVERGROW IP EC-funded project;
- HAGGLE IP EC-funded project;
- NetRefound FET Open project.

She acted as a referee for the foreign PhD Preliminary exam of Ramses Morales, University of Illinois, Urbana Champain, USA.

9.2. Academic teaching

There is a strong teaching activity in the ASAP project team as three of the permanent members are Professor or Assistant Professor.

Anne-Marie Kermarrec and Michel Raynal are each responsible of a Master's courses (University of Rennes 1 and ENS Cachan, Brittany extension) entitled respectively "peer-to-peer systems and applications (PAP)" and "Foundations of Distributed Systems". The teaching in the PAP module is shared with Gabriel Antoniu from the PARIS project-team.

Fabrice le Fessant is an associate professor at Ecole Polytechnique.

Achour Mostefaoui is responsible of a Master's course (University of Bougie, Algeria) entitled "Distributed Algorithms".

Marin Bertier is responsible of the 5th year of the Engineer school INSA Rennes and responsible of a Master's course entitled "Operating System"(INSA)

On November 2007, Aline Carneiro Viana gave 12 hours of Master's courses (M2R) at the University of Paris-Sud.

In addition five Ph.D students, François Bonnet, Yann Busnel, Etienne Rivière, Corentin Travers and Gilles Trédan are Teaching Assistants (*moniteurs*).

9.3. Conferences, seminars, and invitations

Only the events not listed elsewhere are listed below.

Invited Talks

- A.-M. Kermarrec has been an invited speaker at the MINEMA Winter School on Middleware for Mobile Computing in February 2007 and at the DYNAMO Workshop in September 2007 (Action of the European Cost Program on "Foundations and Algorithms for Dynamic Networks").
- Michel Raynal has been an invited speader at the *6th IEEE International Symposium on Network Computing and Applications (NCA'07)*, in July 2007 and At the 21th Int'l Symposium on Distributed Computing (DISC'07) in September 2007.

Seminars:

- A.-M. Kermarrec has been invited to give a seminar at INRIAGrenoble - Rhône-Alpes in February 2007 (invitation from the MESCAL project-team), INRIASophia Antipolis - Méditerranée in March 2007 (invitation from the GEOMETRICA project-team), INRIABordeaux - Sud-Ouest (invitation from the PHOENIX project-team) in September 2007.
- M. Raynal has been an invited to give a seminar at the INRIA Sophia Antipolis - Méditerranée in January 2007; at the Microsoft Beijing in October 2007 (The renaming problem); at the University of Texas at Austin, June 2007, at the Hong-Kong Polytechnic University, March 2007, and at the Wien Technical University April 2007.
- M Bertier has an invited paper at the LADIS workshop, Haifa, Israel, March 2007.

Lectures: Michel Raynal has been invited to give lectures at UNAM, Mexico, “Lectures on distributed computing” in October 2007.

9.4. Administrative responsibilities

A.-M. Kermarrec is an elected member of the INRIAEvaluation Committee since September 2005.

She was a member of the 2007 INRIASelection Committee for the Junior Researcher permanent positions (CR2) at the INRIARennes - Bretagne Atlantique, Grenoble - Rhône Alpes and Sophia Antipolis - Méditerranée Research Units.

She was a member of the 2007 INRIASelection Committee for the Senior Researcher permanent positions (DR2).

She is a member of the working group *Actions initiatives* of the INRIA*Conseil d'Orientation Scientifique et Technologique of INRIA*

10. Bibliography

Major publications by the team in recent years

[1] 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", vol. 56, n^{o} 8, 2007, p. 1055–1070.

[2] R. FRIEDMAN, D. GAVIDIA, L. RODRIGUES, S. VOULGARIS, A. C. VIANA. *Gossiping on MANETs: the Beauty and the Beast*, in "Operating Systems Review", vol. 41, n^{o} 5, October 2007, p. 67–74.

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

[4] S. GORENDER, R. MACEDO, M. RAYNAL. *An adaptive programming model for fault-tolerant distributed computing*, in "IEEE Transactions on Dependable and Secure Computing", To appear, 2007.

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

[6] 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.

[7] A. MOSTÉFAOUI, 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", vol. 61, n$^{\mathrm{o}}$ 1, 2007, p. 125–129.

Year Publications

Doctoral dissertations and Habilitation theses

[8] V. GRAMOLI. *Distributed Shared Memory for Large-Scale Dynamic Systems*, Ph. D. Thesis, Université of Rennes 1, November 2007.

[9] E. LE MERRER. *Decentralised Protocols for Large-Scale Logical Network Management*, Ph. D. Thesis, Université of Rennes 1, November 2007.

[10] É. RIVIÈRE. *Réseaux logiques collaboratifs pour la recherche décentralisée dans les systèmes à large-échelle*, Ph. D. Thesis, Université de Rennes 1, 2007.

[11] É. RIVIÈRE. *Collaborative Overlay Networks for Decentralized Search in Large-Scale Distributed Systems*, Ph. D. Thesis, Université of Rennes 1, November 2007.

[12] C. TRAVERS. *Weak Synchronisation in Asynchronous Distributed Systems*, Ph. D. Thesis, Université of Rennes 1, November 2007.

Articles in refereed journals and book chapters

[13] 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", vol. 56, n$^{\mathrm{o}}$ 8, 2007, p. 1055–1070.

[14] P. COSTA, V. GRAMOLI, M. JELASITY, G. P. JESI, E. LE MERRER, A. MONTRESOR, L. QUERZONI. *Exploring the Interdisciplinary Connections of Gossip-based Systems*, in "Operating Systems Review - Special topic: Gossip-Based Networking", vol. 41, n$^{\mathrm{o}}$ 4, oct 2007.

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

[16] R. FRIEDMAN, A. MOSTÉFAOUI, M. RAYNAL. *On the Respective Power of Diamond P and Diamond S to Solve One-Shot Agreement Problems*, in "IEEE Transactions on Parallel and Distributed Systems", vol. 18, n$^{\mathrm{o}}$ 5, 2007, p. 589–597.

[17] 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", vol. 20, n$^{\mathrm{o}}$ 4, November 2007.

[18] R. GUERRAOUI, M. RAYNAL. *The Alpha of indulgent consensus*, in "The Computer Journal", vol. 50, n$^{\mathrm{o}}$ 1, 2007, p. 53–67.

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

[20] A.-M. KERMARREC, M. VAN STEEN. *Gossiping in Distributed Systems.*, in "ACM Operating System Review", vol. 21, n^{o} 5, October 2007.

[21] A. MOSTÉFAOUI, 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", vol. 61, n^{o} 1, 2007, p. 125–129.

Publications in Conferences and Workshops

[22] K. ALEKEISH, P. EZHILCHELVAN, F. BONNET. *Consensus When Coverage Cannot Be Complete*, in "Workshop on Dependable Application Support for Self-Organizing Networks co-located, Edinburgh, UK", Jun 2007.

[23] O. BEAUMONT, A.-M. KERMARREC, L. MARCHAL, E. RIVIÈRE. *VoroNet: a scalable object network based on Voronoi Tessellations.*, in "Proceedings of 21st IEEE International Parallel & Distributed Processing Symposium (IPDPS)., Long Beach, CA, USA", March 2007.

[24] O. BEAUMONT, A.-M. KERMARREC, E. RIVIÈRE. *Peer to peer multidimensional overlays: Approximating complex structures*, in "In OPODIS, 11th International conference on principles of distributed systems, Guadeloupe, France", December 2007.

[25] F. BONNET, A.-M. KERMARREC, M. RAYNAL. *Epidemic-based Small-World Networks*, in "Proceedings of Workshop Locality, Portland, OR, USA", Aug 2007.

[26] F. BONNET, A.-M. KERMARREC, M. RAYNAL. *Small world networks: From theoretical bounds to pratcical systems.*, in "In OPODIS, 11th International conference on principles of distributed systems, Guadeloupe, France", December 2007.

[27] Y. BUSNEL, A.-M. KERMARREC. *ProxSem: Interest-based Proximity Measure to Improve Search Efficiency in P2P Systems*, in "4th European Conference on Universal Multiservice Networks (ECUMN2007), Toulouse, France", Feb 2007.

[28] A. FERNANDEZ, V. GRAMOLI, E. JIMENEZ, A.-M. KERMARREC, M. RAYNAL. *Distributed slicing in dynamic systems.*, in "Proceedings of ICDCS 2007, Toronto, Canada", June 2007.

[29] A. FERNANDEZ, E. JIMENEZ, M. RAYNAL. *Electing an eventual leader in an asynchronous shared memory system*, in "37th Int'l IEEE Conference on Dependable Systems and Networks (DSN'07)", jun 2007, p. 399–408.

[30] A. FERNANDEZ, E. JIMENEZ, M. RAYNAL, G. TREDAN. *A Timing Assumption and a t-Resilient Protocol for Implementing an Eventual Leader Service in Asynchronous Shared Memory Systems*, in "Proc. 10th Int'l IEEE Symposium on Objects and Component-oriented Real-time Computing (ISORC 2007)", may 2007, p. 71–78.

[31] E. GAFNI, M. RAYNAL, C. TRAVERS. *Test&set, adaptive renaming and set agreement: a guided visit to asynchronous computability*, in "26th IEEE Symposium on Reliable Distributed Systems (SRDS'07)", oct 2007, p. 93–102.

[32] V. GRAMOLI, E. ANCEAUME, A. VIRGILLITO. *SQUARE: Scalable Quorum-Based Atomic Memory with Local Reconfiguration*, in "Proceedings of the 22nd ACM Symposium on Applied Computing (SAC'07)", ACM Press, mar 2007, p. 574–579.

[33] V. GRAMOLI. *Mémoires partagées distribuées pour systèmes dynamiques à grande échelle*, in "8éme Journées Doctorales en Informatique et Réseaux (JDIR'07)", IEEE France Section, jan 2007, p. 153–160.

[34] V. GRAMOLI, A.-M. KERMARREC, A. MOSTÉFAOUI, M. RAYNAL, B. SERICOLA. *Persistance de noyau dans les systemes dynamiques a grande echelle*, in "9ème rencontres francophones sur les aspects algorithmiques de télécommunications (AlgoTel'07)", may 2007.

[35] V. GRAMOLI, M. RAYNAL. *Timed Quorum Systems for Large-Scale and Dynamic Environments*, in "Proceedings of the 11th International Conference On Principles Of Distributed Systems (OPODIS'07)", LNCS, vol. 4878, Springer-Verlag, dec 2007, p. 429–442.

[36] R. GUERRAOUI, M. RAYNAL. *A universal construction for wait-free objects*, in "Proc. ARES 2007 Workshop on Foundations of Fault-tolerant Distributed Computing (FOFDC 2007)", apr 2007, p. 959–966.

[37] C. IGNAT, G. OSTER, P. MOLLI, M. CART, J. FERRIÉ, A.-M. KERMARREC, P. SUTRA, L. BENMOUFFOK, J.-M. BUSCA, M. SHAPIRO, R. GUERRAOUI. *A comparison of optimistic approaches to collaborative editing of Wiki.*, in "In the 3rd International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborativeCom'2007), New-York, USA", November 2007.

[38] F. LE FESSANT. *Gestion de versions de formats avec Camlp4*, in "Dix-huitièmes Journées Francophones des Langages Applicatifs, Aix-les-bains, France", INRIA, 2007.

[39] 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.

[40] A. MOSTÉFAOUI, M. RAYNAL, C. TRAVERS. *From renaming to k-set agreement*, in "14th Colloquium on Structural Information and Communication Complexity (SIROCCO'07)", jun 2007, p. 62–76.

[41] A. MOSTÉFAOUI, G. TREDAN. *Towards the minimal synchrony for byzantine consensus*, in "Brief annoucement, ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC 2007), Portland, OR, USA", Aug 2007.

[42] A. MOSTÉFAOUI. *Towards a Computing Model for Open Distributed Systems*, in "Proc. of the 9th International Conference on Parallel Computing Technologies (PaCT-07)", LNCS, n^{o} 4671, Springer Verlag, September 2007, p. 74-79.

[43] J. RAHMÉ, A. C. VIANA, K. A. AGHA. *Avoiding energy-compromised hotspots in resource-limited wireless networks*, in "IFIP 1st Home Networking Conference (IHN), Paris, France", Dec 2007.

[44] M. RAYNAL, G. TAUBENFELD. *The notion of a timed register and its application to indulgent synchronization*, in "19th ACM Symposium on Parallel Algorithms and Architectures (SPAA'07)", may 2007, p. 200–209.

[45] S. VOULGARIS, E. RIVIÈRE, A.-M. KERMARREC, M. VAN STEEN. *Sub-2-Sub: Self-Organizing Content-Based Publish and Subscribe for Dynamic and Large Scale Collaborative Networks*, in "IPTPS'06: the fifth International Workshop on Peer-to-Peer Systems, Santa Barbara, USA", FEB 2007.

Internal Reports

[46] V. GRAMOLI, M. RAYNAL. *Timed Quorum System for Large-Scale Dynamic Environments*, Technical report, n° 1859, INRIA Research Centre Rennes, July 2007.

References in notes

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

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

[49] K. BIRMAN, M. HAYDEN, O. OZKASAP, Z. XIAO, M. BUDIU, Y. MINSKY. *Bimodal Multicast*, in "ACM Transactions on Computer Systems", vol. 17, n° 2, May 1999, p. 41-88.

[50] A. DEMERS, D. GREENE, C. HAUSER, W. IRISH, J. LARSON. *Epidemic algorithms for replicated database maintenance*, in "Proceedings of the Sixth Annual ACM Symposium on Principles of Distributed Computing (PODC'87), Vancouver, British Columbia, Canada", August 1987, p. 1-12.

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

[52] M. JELASITY, R. GUERRAOUI, A.-M. KERMARREC, M. VAN STEEN. *The Peer Sampling Service: Experimental Evaluation of Unstructured Gossip-Based Implementations*, vol. 52, n° 2, February 2003.

[53] L. LAMPORT. *Time, clocks, and the ordering of events in distributed systems*, in "Communications of the ACM", vol. 21, n° 7, 1978.

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

[55] S. RATNASAMY, P. FRANCIS, M. HANDLEY, R. KARP, S. SHENKER. *A Scalable Content-Addressable Network*, in "Proceedings of ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'01), New York, NY, USA", August 2001, p. 161–172.

[56] 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)", November 2001, p. 329–350.

[57] I. STOICA, R. MORRIS, D. KARGER, M. F. KAASHOEK, H. BALAKRISHNAN. *Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications*, in "Proceedings of the 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'01), San Diego, CA, USA", ACM Press, ACM, August 2001, p. 149–160, http://www.pdos.lcs.mit.edu/.

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