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, Saclay - Île-de-France

THEME COM

Activity Report

2008

Table of contents

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.*

1. Team

Research Scientist

Anne-Marie Kermarrec [DR, INRIA RENNES - BRETAGNE ATLANTIQUE, HdR]
Gérard Le Lann [DR, INRIA RENNES - BRETAGNE ATLANTIQUE (since September 2008), HdR]
Fabrice Le Fessant [CR1, INRIA SACLAY - ILE DE FRANCE SUD]
Aline Carneiro Viana [CR2, INRIA SACLAY - ILE DE FRANCE SUD]

Faculty Member

Achour Mostefaoui [Associate Professor (MdC), University Rennes 1, HdR]
Michel Raynal [Professor (Pr), University Rennes 1, HdR]
Marin Bertier [Assistant Professor (MdC), INSA Rennes]

Technical Staff

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

PhD Student

Kévin Huguenin [ENS CACHAN Grant,MENRT Grant, INRIA RENNES - BRETAGNE ATLANTIQUE]
Damien Imbs [MENRT Grant (since October 2008),INRIA RENNES - BRETAGNE ATLANTIQUE]
Nicolas Le Scouarnec [Cifre Thomson Grant,INRIA RENNES - BRETAGNE ATLANTIQUE]
François Bonnet [MENRT,INRIA RENNES - BRETAGNE ATLANTIQUE]
Xiao Bai [INSA- UTC - Chinese Council,INRIA RENNES - BRETAGNE ATLANTIQUE]
Yann Busnel [MENRT Grant (until November 2008),INRIA RENNES - BRETAGNE ATLANTIQUE]
Vincent Leroy [MENRT Grant,INRIA RENNES - BRETAGNE ATLANTIQUE]
Gilles Trédan [MENRT Grant,INRIA RENNES - BRETAGNE ATLANTIQUE]
Mubashir Rehmani [MENRT Grant (since October 2008),INRIA SACLAY - ILE DE FRANCE SUD]

Post-Doctoral Fellow

Davide Frey [Post-Doc INRIA RENNES - BRETAGNE ATLANTIQUE]
Sathya Peri [Post-Doc INRIA RENNES - BRETAGNE ATLANTIQUE]
Guang Tan [Post-Doc INRIA RENNES - BRETAGNE ATLANTIQUE]
Cigdem Sengul [Post-Doc INRIA SACLAY - ILE DE FRANCE SUD (November 2007-November 2008)]
Golnaz Karbaschi [Post-Doc INRIA SACLAY - ILE DE FRANCE SUD (since September 2008)]

Visiting Scientist

Artur Ziviani [LNCC, Brazil, since September 2008]

Administrative Assistant

Cécile Bouton [INRIA RENNES - BRETAGNE ATLANTIQUE]
Marie-Jeanne Gaffard [INRIA SACLAY - ILE DE FRANCE SUD]

Other

Silvia Bonomi [Internship INRIA RENNES - BRETAGNE ATLANTIQUE (Sep-Feb 2008)]
François Joulaud [Master student UNIVERSITY RENNES 1, February-July 2008]
Hamouma Moumen [PhD candidate April-May 2008, INRIA RENNES - BRETAGNE ATLANTIQUE]
Antonis Papadimitriou [Internship INRIA SACLAY - ILE DE FRANCE SUD (May-July 2008)]
Imen Chakroun [Master student,INRIA SACLAY - ILE DE FRANCE SUD (since October 2008)]
Massimo Vecchio [Internship,INRIA SACLAY - ILE DE FRANCE SUD (since October 2008)]
Mubashir Rehmani [Master student, INRIA SACLAY - ILE DE FRANCE SUD (April-August 2008)]

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*" [51] (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. Two main classes have been designed: (1) structured overlay networks, such as distributed hash tables (DHT), rely on a logical structure and map object keys to overlay nodes. While structured networks were initially more popular among scientists, their exact-match interface limits their flexibility and use for various applications, notably for non-exact information retrieval. (2) 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 this area, our approach is original for three reasons. First, 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. 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), because of the heterogeneity and limitation of resources among participating entities.

Second, we strongly believe in weakly-structured networks and most of our projects rely on epidemic-based unstructured overlays. 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 [50].

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 to develop our own platform in the context of our peer-to-peer collaborative backup platform.

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 [52], [45], [46]. 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 [53], [54], [55] following a specific topology or an unstructured network [49], [56] 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 [48], [47], 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 [47], 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

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

Contact:	Fabrice Le Fessant
Licence:	GPL
Presentation:	Friend-to-Friend Private Peer-to-Peer Client
Status:	under development

Peerple 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 (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 working on adding distributed backups and code modularization.

5.2. Move&Play and MNPlight

Participant: Fabrice Le Fessant.

Contact:	Fabrice Le Fessant
Licence:	Proprietary
Presentation:	Scalable Online File-Sharing System
Status:	under development

MoveNPlay is a web-based system enabling its users to easily access and share their content online. It provides a Web 2.0 interface, and uses online storage from the Amazon S3 service. We are now working on embedding the system for setup boxes of ISP providers and building a scalable architecture for the database, switching from Sqlite to MySql. MoveNPlay is developed in the context of the creation of a startup, who received a ANVAR/OSEO grant in may 2008.

MNPlight is a specialization of Peerple, for the iPhone, developped to interact with MoveNPlay. It provides many functions that are not available natively on the iPhone, such as calendar synchronization, mail search, music upload without iTunes, etc. In particular, it helps to replace the lack of iTunes for Linux users. It is distributed since October 2007, and we are now porting it to the version 2.0 of the iPhone operating system.

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

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

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

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

Fabrice: this requires verification and update

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.

5.4. Java library for effective development of gossip-based applications

Participants: Davide Frey, Anne-Marie Kermarrec.

Contact:	Davide Frey
Licence:	Proprietary
Presentation:	Library for Gossip protocols
Status:	released version 0.1

GossipLib is a library consisting of a set of JAVA classes aimed to facilitate the development of gossip-based application in a large-scale setting. The current version of GossipLib provides the implementation of a peer-sampling protocol, as well as a demo application enabling the visualization of the execution of the protocols. The architecture of GossipLib is designed to facilitate code-reuse. Each gossip-based component may be used as a building block to develop new and more complex protocols.

The new version of GossipLib, currently under development, will extend the library in several directions. First, it will integrate with GossiPeer to manage deployment in a large-scale setting. Second, it will integrate with a simulation engine to enable rapid testing and evaluation of applications developed using the library. Finally, we are incorporating a runtime substrate capable of dynamically managing multiple cooperating gossip-based applications distributed over a large set of network nodes.

5.5. SeNSim simulator and visualisator

Participants: Marin Bertier, Yann Busnel, Gilles Trédan.

Contact:	Marin Bertier
Licence:	Not defined yet
Status:	under development
Presentation:	Simulator for Sensor Networks
URL:	http://sensim.gforge.INRIA.fr/

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. 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.

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.

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. 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.

6.2.2. Distributed slicing

Participant: Anne-Marie Kermarrec.

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 by extending our previous work on the topic. This work has been done in collaboration with Vincent Gramoli (EPFL & University of Neuchatel, Switzerland), Ymir Vigfusson, Ken Birman and Robbert van Renesse (Cornell University). More specifically, we proposed Sliver a fast converging algorithm achieving accurate slicing.

6.2.3. Peer-to-peer Polling without cryptography

Participants: Kévin Huguenin, Anne-Marie Kermarrec.

This work has been done in collaboration with Rachid Guerraoui and Maxime Monod (EPFL, Switzerland). The emergence of social networks provides a framework for polling a community easily by the mean of peer-to-peer techniques. Polling is not as critical as voting as the accuracy on the tally is less important. Yet, it must provide similar properties to electronic voting, such as voters privacy, fairness, probabilistic accuracy, ...The core idea of the work is to build a decentralized protocol without cryptography ensuring this properties with high probability by the mean of peer-to-peer deterrent power : *every action in the protocol may be subject to verification by peers*. Ensuring that any malicious action is detected with probability one or at least close to one, we increase the accuracy of the tally by limiting the proportion of peers misbehaving. Privacy is ensured probabilistically by using peers as proxies for emitting ballots making vote recovery impossible for reasonable proportion of malicious nodes.

6.2.4. Conditions for set agreement

Participants: François Bonnet, Michel Raynal.

We applied the *condition-based* approach to the k-set agreement problem. In a brief annoucement in PODC'08, we introduced the definition of a framework that allows defining conditions suited to the ℓ-set agreement problem. More precisely, a condition is defined as a set of input vectors such that each of its input vectors can be seen as "encoding" ℓ values, namely, the values that can be decided from that vector. A condition is characterized by the parameters t, ℓ, and a parameter denoted d such that the greater $d + \ell$, the least constraining the condition (i.e., it includes more and more input vectors when $d + \ell$ increases, and there is a condition that includes all the input vectors when $d + \ell > t$). The conditions characterized by the triple of parameters t, d and ℓ define the class of conditions denoted $\mathbb{S}_t^{d,\ell}$, $0 \leq d \leq t$, $1 \leq \ell \leq n - 1$. The properties of the sets $\mathbb{S}_t^{d,\ell}$ are investigated, and it is shown that they have a lattice structure.

In ICDCS'08 [13], our contribution is a generic synchronous k-set agreement algorithm based on a condition $C \in \mathbb{S}_t^{d,\ell}$, i.e., a condition suited to the ℓ-set agreement problem, for $\ell \leq k$. This algorithm requires at most $\lfloor \frac{d-1+\ell}{k} \rfloor + 1$ rounds when the input vector belongs to C, and $\lfloor \frac{t}{k} \rfloor + 1$ rounds otherwise. (Interestingly, this algorithm includes as particular cases the classical synchronous k-set agreement algorithm that requires $\lfloor \frac{t}{k} \rfloor + 1$ rounds (case $d = t$ and $\ell = 1$), and the synchronous consensus condition-based algorithm that terminates in $d + 1$ rounds when the input vector belongs to the condition, and in $t + 1$ rounds otherwise (case $k = \ell = 1$).)

6.2.5. Anonymous graph exploration without collision by mobile robots

Participants: François Bonnet, Michel Raynal.

We studied autonomous mobile robots moving on a finite anonymous graph and focuse on the *Constrained Perpetual Graph Exploration* problem ($CPGE$). That problem requires each robot to perpetually visit all the vertices of the graph, in such a way that no vertex hosts more than one robot at a time, and each edge is traversed by at most one robot at a time. This work has been done in cooperation with Roberto Baldoni professor in Roma and Alessia Milani post-doc in Madrid. François Bonnet spent one week in Roma to work on this subject and he also presented this work during a seminar ALPAGE (February'08) and as an invited speaker in LIP6 (November'08).

In an Information Processing Letters paper [8], we provided an upper bound k on the number of robots that can be placed in the graph while keeping $CPGE$ solvability. In a short paper in OPODIS'08, The second contribution restricts. We propose algorithms that solve the $CPGE$ problem up to a certain number k of robots depending on a given parameter ρ that captures the view of each robot, on partial anonymous grid (a finite grid with possibly missing vertices or edges).

6.2.6. FADAS: Formalisms and Algorithms for Resilient Services Design in Ambient Systems

Participants: Marin Bertier, François Bonnet, Michel Raynal.

We focus on the problem of providing a reliable, geographically localised storage service. This problem can be abstracted as providing a virtual shared register, i.e. an object equipped with read and write functionalities, attached to a particular location and maintained by a set of mobile nodes. Such an abstraction can be seen as a basic building block to be used for the implementation of higher level services like a localised warning system or a localised black board.

The preliminary results of our work are published as short papers in EDCC'08, DEBS'08, and OPODIS'08 conferences. We specified and implemented such a shared object for non concurrent writes and multiple readers, then we describe a possible extension to concurrent writes. We also provide insights on how this kind of objects could be useful to program location-aware applications. This work is a collaboration between people from Laas in Toulouse, people from La Sapienza in Roma, and people from IRISA in Rennes as a mini-project in the European Network of Excellence ReSIST. In this context, François spent few days in Roma, and in Toulouse twice.

6.2.7. On the Impact of the Mobility on Convergence Speed of Population Protocols

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

Population protocols provide theoretical foundations for mobile tiny device networks in which global behavior emerges from a set of simple interactions between anonymous agents. The works in this area mostly focus on studying the computational power of the model. Results hold as long as a fair scheduler, ensuring that all reachable system states are endlessly reached, governs the interactions between nodes. This assumption is crucial to ensure that the protocols eventually converge.

We studied for the first time the impact of the agents mobility model on the convergence speed of population protocols. We perform our study by considering several mobility models traditionally used in the ad-hoc network community. We propose an augmented population protocol model where each edge of the interaction graph is weighted, representing the probability of two agents to interact. This models the behaviour of the scheduler with respect to various mobility models. We empirically show that mobility models do have a significant impact on the convergence speed of the protocols. In fact, we observe that the uniform distribution always provides the best convergence time. Such a model is representative of the well-known random-way point model used to evaluate most of mobile ad-hoc network protocols. We then formally proved that a uniform distribution of weights provides the lowest bound of average convergence speed for any population protocol.

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. GOSSPLE : A radically new approach to navigating the digital information universe

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

This work started on September 1st 2008 and is the topic of the ERC Starting Grant led by Anne-Marie Kermarrec. Traditional search engines are performing extremely well but do hardly encompass alternative and very dynamic sources of information such as blogs, p2p file sharing system instant messaging as well as some modern forms of content distribution frameworks. This is mainly due to their lack of adaptivity to dynamics and their not taking into account correlations between contents and users preferences. They are also limited by their reliance upon centralized indexing. Typically, corporate pages are visited frequently while individual information may be visited rarely: the individual is at a disadvantage. This reveals a striking evidence that complementary and novel fully decentralized alternatives to traditional search engines are now required to capture the dynamic, collaborative and heterogeneous nature of the digital universe as well as leverage individual preferences and social affinities. At the heart of GOSSPLE lies the idea that leveraging the virtual communities of users and the social dimension of the network in a dynamic manner is the key towards reaching the many areas of the Internet applications that are not covered by current search engines. GOSSPLE stems from the observation that indexing the data is not really the main issue, collaborative and dynamic navigation is. GOSSPLE is an autonomic, completely decentralized self-organized network, in which participants get connected to the virtual sub-network, i.e., an overlay network, matching the operations they are performing.

We initiated GOSSPLE in the context of user-centric search. In this context, we worked on the metrics to assess the distance between users in arge-scale systems. Collaborative filtering algorithms are information retrieval techniques used to detect interest patterns of users for items. They are useful to predict the interest of a user for a given item. Those technique are already deployed in many applications, but they only work in a centralized context and are therefore limited by storage and calculation time constraints. We developed fully decentralized algorithms to automatically group users whose interests are similar and provide each user with a list of items that match his profile. We focused both on the quality of the proposed items and their diversity. Our approach was validated through simulations on data from the music community website lastFM and eDonkey file sharing logs. We are currently investigating more complex tag-based systems. Vivien Quéma (SARDES) and Rachid Gerraoui (EPFL) are currently collaborating with us on this topic.

6.3.2. Efficient Geographic routing in large-scale networks.

Participants: Marin Bertier, Anne-Marie Kermarrec, Guang Tan.

Geographic routing is a hot topic in the wireless sensor network area, it consists to route a message from any nodes to the closer node of the destination coordinates. In this topic we improve previous works particularly for complex topology. Our first contribution was to propose a routing protocol which achieves optimal routing performance with a message and memory overhead per node independent of the network size.

M. Bertier served in organization committee for the following conferences: Algotel 2008 Rencontres Francophones sur les aspects Algorithmiques de Télécommunications, Saint-Malo, France, May 2008. DEBS'08 2nd Int'l Conference on Distributed Event-Based Systems, Rome, Italy, Jul 2008. He served also in program committes of the following conferences:

Algotel 2008 Rencontres Francophones sur les aspects Algorithmiques de Télécommunications, Saint-Malo, France, May 2008. CFIP'08: Colloque francophone sur l'ingénierie des protocoles, Arc, France, 2008. CFSE'08: Conférence Française en SystÃ¨mes d'Exploitation, Fribourg, 2008 IWSOS 2008, the third International Workshop on Self-Organizing Systems, Vienna, Austria, December 2008

6.3.3. Gossip-based overlay networks

6.3.3.1. Bridging the Gap between Population and Gossip-based Protocols

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

Gossip-based protocols are simple, robust and scalable and have been consistently applied in many distributed, mostly wired, settings. Most validation in this area has been so far empirical and there is a clear lack of a theoretical counterpart clearly defining what can and cannot be computed with gossip-based protocols. Population protocols, on the other hand, provide a clear and formal model for mobile sensor networks, capturing their power and limitations.

We established a correlation between population and gossip-based protocols. Studying the equivalence between them, we propose a classification of gossip-based protocols, based on the nature of the underlying peer sampling service. First, we show that the class of gossip protocols, where each node relies on an arbitrary sample, is equivalent to population protocols. Second, we show that gossip-based protocols, relying on a more powerful peer sampling providing peers using a clearly identified set of other peers, are equivalent to community protocols, a variant of population protocols. Leveraging the resemblances between these areas enables to provide a theoretical framework for distributed systems where global behaviours emerge from a set of local interactions, both in wired and wireless settings.

6.3.3.2. PULP: Push Pull Gossip protocols

Participant: Anne-Marie Kermarrec.

This work has been done in collaboration with Etienne Rivière and Pascal felber (university of neuchatel, Switzerland) and Spyros voulagris (ETH Zurich). Gossip-based protocols are well known to provide a simple, scalable and extremely robust way to disseminate messages in large-scale systems. We proposed a generic efficient push-pull dissemination protocol, PULP, which combines the best of both worlds. PULP exploits the reliability and efficiency of push approaches, while limiting redundant messages and therefore imposing a low overhead, as pull protocols do. PULP relies on the dissemination of a stream of messages to achieve this: by pushing some messages, PULP enables an efficient pulling of other messages, which in turn help with the dissemination of the stream. We have deployed PULP both on a cluster and on PlanetLab. Our results demonstrate that PULP achieves an appealing trade-off between robustness, redundancy, and delays.

6.3.3.3. Incentive-compatible peer-to-peer Video-on-Demand

Participants: Kévin Huguenin, Anne-Marie Kermarrec.

This work has been done in collaboration with Maarten Van Steen and Vivek Rai (Vrije Universiteit Amsterdam). Video-on-demand (VoD) is increasingly attractive internet application allowing a client to begins to play a movie stored on a server almost instantaneously. The missing pieces of the video are downloaded in a sequential order while playing the downloaded pieces. Peer-to-peer swarming has proved to be a very efficient paradigm for file distribution even with a very low proportion of seeds provided that all the peers actually contribute to the system by uploading pieces to others. This problem is critical in VoD applications since peers must achieve their deadlines, and thus rely on the other peers, to provide a smooth playback to the user. The incentives used by Bit Torrent, referred as Tit for tat are not usable in VoD downloads due to the sequential nature of VoD downloads and since any two peers may not have mutual interest.

Our work consist in building an efficient Incentive-compatible peer-to-peer Video-on-Demand system by imposing a loose structure on the peer sets : grouping nodes whose positions in the video are close in groups and build a linked list of groups to ensure *feeding* groups by more advanced ones in exchange of most advanced pieces injected in less advanced groups. Using an appropriate pieces seeding/feeding strategy the full swarm comes down to several very efficient independent swarms fed by each others, the seed being appealed only by the most advanced group. The structured VoD swarming protocol outperforms existing protocols based on random graphs in terms of throughput, goodput (i.e., playback rate) and memory usage (by keeping a limited memory of past pieces instead of storing the entire video).

6.3.3.4. 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. Following our theoretical analysis in 2007 in which we refined the routing complexity measure for small-world networks, we are now working on the practical evaluation of a Kleinberg based peer sampling. More specfically, 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. We showed that such a biased peer sampling impact only a little the properties of the resulting graph such as the average path length, the clustering coefficient and the in-degree distribution.

6.3.3.5. Heterogeneous gossip

Participants: Davide Frey, Anne-Marie Kermarrec.

This work has been done in collaboration with Vivien Quéma (CNRS Grenoble), Maxime Monod and rachid Guerraoui (EPFL). Gossip-based information dissemination protocols are considered easy to deploy, scalable and resilient to network dynamics. 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 network capabilities such as bandwidth. In practice, a blind load-balancing strategy might significantly hamper the performance of the gossip dissemination.

These observation, collected in a white paper presented at LADIS 2008 as an invited contribution [12], have led to the development of HEAP, *HEterogeneity-Aware gossip Protocol*, where nodes dynamically adapt their contribution to the gossip dissemination according to their bandwidth capabilities. Using a continuous, itself gossip-based, approximation of relative bandwidth capabilities, HEAP dynamically leverages the most capable nodes by increasing their fanout, while decreasing by the same proportion those of less capable nodes. HEAP preserves the simplicity and proactive (churn adaptation) nature of gossip, while significantly improving its effectiveness. We extensively evaluated HEAP in the context of a video streaming application on a 270 PlanetLab node testbed. Our results show that HEAP significantly improves the perceived quality of the streaming over standard gossip protocols, especially when the stream rate is close to the average available bandwidth.

6.3.3.6. Combining gossip and fountain codes for fast dissemination

Participants: Anne-Marie Kermarrec, Nicolas Le Scouarnec.

Gossip-based dissemination reaches all nodes with high probability. However, some peers may not receive all the data. In order to enhance the performance of gossip-based dissemination protocol, fountain codes can be used to recover the missing data. We propose a protocol that combine a gossip-based dissemination protocol and fountain codes. The gossip protocol builds a structured split-graph overlay to split the peers between encoders or forwarders. Forwarders become encoders as soon as they have received the whole content so they can start encoding. In order to benefit even further from encoders, we deliberately bias the dissemination process through the gossip protocol so that forwarders become earlier potential encoders. We implemented this protocol on the PlanetLab testbed. By simulation in PeerSim, we assess that this protocol outperform a simple protocol that do not introduce a bias or do not update the overlay to match the peers'role.

6.3.3.7. On the Fly DHT based on Gossip Protocols

Participants: Marin Bertier, François Bonnet, Anne-Marie Kerrmarec, Vincent Leroy, Sathya Peri, Michel Raynal.

DHTs and RPS are two very important services used by many P2P applications. DHT based systems provide a efficient lookup functionality while peer sampling services enable to build and maintain connectivity in the presence of high churn and offer a lot of flexibility. In addition many applications, in particular search applications, may want to combine them to leverage the benefits of both approaches. We propose to integrate a p2p structured overlay and gossip-based overlay sampling service to build a DHT on the fly. The small-world nature of DHTs based systems is captured by the use of two cohabiting gossip-based protocols: SRLS maintains a set of short links (numerically speaking) and LRLS selects long-range links.

Our approach has several advantages: It can start from arbitrarily connected overlay and build the system on the fly as nodes gossip. It is very simple and unlike other DHT protocols has no special mechanism to handle churn and even extreme cases such as catastrophic failures. The set of long-range which are changing continuously provide a very useful feature of distributing the routing load evenly across the system. The indegree of the nodes are well balanced.

6.3.4. Failure-Tolerant Overlay Trees for Large-Scale Dynamic Networks

Participant: Davide Frey.

This work has been done in collaboration with Amy Murphy (Trento, Italy). Trees are fundamental structures for data dissemination in large-scale network scenarios. However, their inherent fragility has led researchers to rely on more redundant mesh topologies in the presence of churn or other highly dynamic settings. In this work [18], we identified a novel protocol that directly and efficiently maintains a tree overlay in the presence of churn. The protocol simultaneously achieves other beneficial properties such as limiting the maximum node degree, minimizing the extent of the tree topology changes resulting from failures, and limiting the number of nodes affected by each topology change. Applicability to a range of distributed applications is discussed and results are evaluated through extensive simulation and a PlanetLab deployment.

6.3.5. Peer-to-peer Collaborative back-up

Participants: Imen Chakroun, Anne-Marie Kermarrec, Fabrice Le Fessant.

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

6.3.6. Monitoring overlay networks

6.3.6.1. A distributed churn measurement method

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

Many of distributed systems are dynamic in that participants may leave or (re-)join the system unpredictably. The key parameter of such systems is called the *churn*, i.e. the dynamism intensity. To face dynamism, existing applications traditionally use costly recovery mechanisms that overestimate the churn. As a consequence, the communication cost employed to recover despite unbounded failures prevents these applications from scaling as well as they could. To avoid the need for upper bounds on churn rate, decentralized estimation schemes provide a way for every system node to self adapt to network characteristics; to provide an estimation, recent approaches assume an underlying structured network (*DHT*) and the fact that churn events are uniformly distributed all over the network. Unfortunately, this strongly restricts their use.

We present an algorithm that is, as far as we know, the first distributed method for monitoring churn in arbitrary networks, with also arbitrary node departure patterns. This method relies on a constant number of message exchanges between each node and its neighbors, and thus scales well. It 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. Results have been published in 2008 in conferences CFSE [22] and PODC [21].

6.3.6.2. Distributed betweenness centrality through a perpetual random walk

Participants: Anne-Marie Kermarrec, Erwan Le Merrer, Gilles Trédan.

This work is done in collaboration with Bruno Séricola (DYONISOS). We propose a distributed algorithm to assess the connectivity quality of a network, be it physical or logical. In large complex networks, some nodes may play a vital role due to their position (*e.g.* for routing or network reliability). Assessing global properties of a graph, as importance of nodes, usually involves lots of communications; doing so while keeping the overhead low is an open challenge. To that end, *centrality* notions have been introduced by researchers to rank the nodes as a function of their importance in the topology. Some of these techniques are based on computing the ratio of shortest-paths that pass through any graph node. This approach has a limitation as nodes “close” from the shortest-paths do not get a better score than any other arbitrary ones. To avoid this drawback, physician Newman proposed a centralized measure of *betweenness centrality* based on random walks: counting for each node the number of visits of a random walk travelling from a node i to a target node j, and then averaging this value by all graph source/target pairs. Yet this approach relies on the full knowledge of the graph for each system node, as the random walk target node should be known by advance; this is not an option in large-scale networks. We propose a distributed solution that relies on a single random walk travelling in the graph; each node only needs to be aware of its topological neighbors to forward the walk. A preliminary paper has been published at DISC conference in 2008 [24].

6.3.6.3. Measuring Availability in Peer-to-Peer Systems

Participants: Anne-Marie Kermarrec, Fabrice Le Fessant, Cigdem Sengul.

Tracking peer availability in a peer to peer network is of utmost importance for many collaborative applications. For instance, such information is invaluable for identifying the most stable peers or group of peers with similar uptime characteristics. However, as many applications tend to reward the most stable peers, there is a clear incentive for peers to try to appear more available than their real availability. We developed a scalable and lightweight protocol that enables nodes to measure the peer availability in the presence of such selfish peers. In our protocol, which is called Pacemaker [41], each peer is in charge of maintaining proofs of its own availability over time by collecting pulses disseminated by a trusted entity using asymmetric cryptographic signatures. Essentially, using these pulses, peers gain the ability to challenge other peers and verify that their real uptime matches the advertised one. Simulation results show that our protocol provides accurate availability measures even in the presence of selfish peers. Furthermore, our results are verified by experiments in Planetlab, which also illustrates the deployability of Pacemaker in real networks.

6.4. Peer-to-peer wireless autonomic networked systems

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

6.4.1. Self-Organizing Object Deployment

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

We worked on the design, correctness, and analysis of SONDe, a simple fully decentralized object deployment algorithm for highly requested systems. Given an object (service or data), SONDe provides a node with a constant upper bound (h) on the number of logical hops to access an object holder (provider), thus making tunable and predictable the communication latency between a node and any provider. In addition, SONDe is able to dynamically adapt the number of providers to reflect load variations experienced in localized portions of the system. Each node individually decides to be a provider, based on the observation of its h-hops neighborhood. We show theoretically that SONDe self-stabilizes and provides an independent-dominating set of providers. Finally simulation results, conducted over different network topologies, demonstrate the efficiency of the approach and confirm the theoretical analysis. This research has been published at EDCC conference in 2008 [23].

6.4.2. Tracking moving objects

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

This work has been done in collaboration with Roberto Baldoni and Leonardo Querzoni (University La sapienza, Rome). We consider a set of anonymous moving objects to be tracked in a binary sensor network, we studied the problem of associating deterministically a track revealed by the sensor network with the trajectory of an unique anonymous object, namely the *Multiple Object Tracking and Identification* (MOTI) problem. In our model, the network is represented by a sparse connected graph where each vertex represents a binary sensor and there is an edge between two sensors if an object can pass from one sensed region to another one without activating any other remaining sensor. The difficulty of MOTI lies in the fact that trajectories of two or more objects can be so close (track merging) that the corresponding tracks on the sensor network can no longer be distinguished, thus confusing the deterministic association between an object trajectory and a track.

We first showed that MOTI cannot be solved on a general graph of ideal binary sensors even by an omniscient external observer if all the objects can freely move on the graph. Then, we restricted the graph, on the object movements or both, to make MOTI problem always solvable. In the case of absence of the omniscient observer, we showed that our results can lead to the definition of distributed algorithms that are able to detect when the system is in a state where MOTI becomes unsolvable.

6.4.3. Adaptive data forwarding in delay tolerant networks

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

We conducted studies in information dissemination in mobile networks where a contemporaneous path may never exist between two nodes (a source and a destination) in the network, known as delay-tolerant networks (DTNs). These networks are further challenged by strict resource constraints (e.g., memory, CPU and energy limitations). Our goal in this domain was to support intelligent and adaptive forwarding, which allows a good trade-off between reliability and resource-efficiency. In particular, the main challenge for adaptive mechanisms in DTNs is to adjust forwarding decisions on the fly and still ensure reachability, even when the network is not always connected and disconnections are hard to predict. We then design a new protocol, called Seeker, which empowers nodes with the ability to estimate *favorable contact opportunities* by taking advantage of any information they can locally infer. Hence, nodes are able to adapt and self-organize in dynamic environments with minimal control overhead. This work has been performed in collaboration with the researcher Roy Friedman (from Technion - Israel Institute of Technology) and is currently in submission. The performed research work was also inserted in the context of the RNRT SVP project, which has finished in September 2008. Related to this described research a study about: data dissemination using gossip algorithms in wireless networks was published in the ACM SIGOPS Operating Systems Review (ACM OSR) and routing in wireless self-organizing networks was published in a chapter of the book Adaptation and Cross Layer Design in Wireless Networks, Taylor and Francis Group [33].

6.4.4. Self-management

6.4.4.1. Self-deployment for free by routing

Participants: Kévin Huguenin, Anne-Marie Kermarrec.

This work has been done in collaboration with Eric Fleury (ENS LYON) Most Wireless Sensor Networks (WSN) applications require connected-coverage of the monitored zone. We consider an application where external entities such as human beings ask the network (through any of its sensor) to sense at a given location in the Region of Interest (ROI) and we call the fulfilment of such a request routing. Using mobile sensors, we propose a routing algorithm independent from the communication devices (i.e., which does not assume a disc model for the communication graph) which provides, in addition to satisfying sensing requests, transparently self-deployment of the network in the ROI. The sensors are not aware of the ROI but only satisfy request either by moving, forwarding or fulfilling sensing requests without any knowledge of the ROI (size, shape, boundaries,...). We prove formally that the network converges to a triangular lattice providing greedy-connected coverage of the region of interest (i.e., once the system has converge no move is needed to fulfil a request anymore) as the number of requests increases. The triangular lattice requires a number of nodes within a constant factor of the optimal connected-coverage deployment. We provide simulations results which confirm the theoretical study and demonstrate the efficiency of the protocol. In addition, we consider practical matters making the algorithm usable in real-life environment. This work has been done in collaboration with Eric Fleury (ENS Lyon / INRIA Rhônes-Alpes).

6.4.4.2. Virtual coordinates for autonomous networked system

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

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 was an autonomous system able not only to be adaptable to environment changing conditions, but also, that provides variety to network functionalities. Thus, we have designed 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 have only exploited local connectivity information and per-neighbor communication. This work has been performed in collaboration with the researcher Roy Friedman (from Technion - Israel Institute of Technology) and is currently in submission. The research related to this subject has been published: at the PODC 2008 (as a brief announcement)[19] and in CFIP 2008 [31] conferences, at the lettre bimestrielle Réeseaux et Télécoms [11], and recently at the ICDCN conference [20] to be held in January 2009.

6.4.4.3. Data organization in wireless sensor networks

Participants: Mubashir Rehmani, Aline Carneiro Viana.

The research activities in this domain concern 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.

We have been working on how to make the sensed monitored data available to the mobile sink in a robust, adaptive, and efficient way. In this context, We have proposed, in collaboration with the researchers Artur Ziviani (from the National Laboratory for Scientific Computing (LNCC), Brazil) and Roy Friedman (from the Technion - Israel Institute of Technology) an efficient data dissemination approach (in terms of overhead and representativeness) to allow a mobile sink to gather a representative view of the monitored region covered by n sensor nodes by visiting **any** m nodes, where $m << n$. This work is currently in submission. This research also constituted the M2R internship subject of Mubashir Rehmani, student of SUPELEC under the supervision of Aline C. Viana. Since September 2008, Mubashir is a Phd student at the UPMC Univ Paris 06, under the supervision of Aline C. Viana and the supervision of the Prof. Serge Fdida. Related to this research a study about data aggregation in wireless sensor networks was also published as a chapter of the book Ad Hoc Networks: New Research, Nova Science Publisher, Inc. [32].

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 the PhD student Joseph Rahme both from the LRI/Universite Paris-Sud, our first contribution in this context was the proposal of cost functions to choose energy efficient routes. The originality of the proposed cost functions lies in their completely decentralized and adaptive behavior in considering energy consumption, remaining energy of nodes, and the number of transmissions a node can make before its energy depletion. The gotten results were published in the Springer Annals of Telecommunication Journal [9] and in the IFIP IHN 2007 conference.

6.4.4.5. Target coverage in wireless sensor networks

Participant: Aline Carneiro Viana.

A 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. 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. To answer this issue, we proposed a mobility strategy that consists in making nodes follow a Hilbert space-filling curve and use opportunistic contacts to reduce the data delivery's delay. The first results on this work were published at the ACM HeterSanet 2008 workshop [16].

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. Byzantine fault tolerant broadcast in wireless sensor networks

Participants: Marin Bertier, Anne-Marie Kermarrec, Guang Tan.

We studies the message efficiency problem of Byzantine fault tolerant broadcast in multi-hop wireless networks. We tries to find the most message or energy efficient ways to accomplish a reliable broadcast despite malicious nodes that may alter the message or cause collisions. We show that the network's ability to tolerate malicious node depends on nodes' message bound, the malicious node density and the topology form by the nodes which have already receive the message. Initial results were presented as short paper at DISC 2008.

7. Contracts and Grants with Industry

7.1. France Telecom

Participant: Anne-Marie Kermarrec.

From October 2004 to november 2007, we had a collaboration with France Télécom R&D, Lannion on applying peer-to-peer techniques to telecom operator frameworks. More specifically, in this area, we worked on timely dissemination of voice over IP and a reliable and distributed telecom infrastructure. In this context, Anne-Marie Kermarrec acted as the PhD advisor of Erwan le Merrer (thesis defended November 2007).

7.2. Thomson

Participants: Anne-Marie Kermarrec, Nicolas Le Scouarnec.

From november 2007 to november 2010, we have a collaboration with Thomson R&D France on providing coding and incentives mechanisms for audio/video secure content distribution over Internet using the peer to peer paradigm. In this context, Anne-Marie Kermarrec acts as the PhD advisor of Nicolas Le Scouarnec.

8. Other Grants and Activities

8.1. National grants

8.1.1. ACI MD 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: (1) large-scale distributed platform modeling, (2) overlay network topologies, (3) scheduling for regular parallel applications and (4) 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 addressed 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 was 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 adaptive data forwarding in delay tolerant networks took place in this context. This project has officially been concluded in September 2008.

8.1.3. Rnrt project SensLAb

Participants: Marin Bertier, Antoine Boutet, 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 Malisse

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

This is a collaboration project between ASAP and ARES INRIAGRENOBLE - RHÔNE-ALPES project team, and EPFL Lausanne (Rachid Guerraoui) school. This project has officially been concluded in October 2008. The goal of this project was to explore the impact of malfunctions and misbehavior (malicious) sensors in a large wide sensor networks. In this context, the members propose several algorithm in particular a fire dection application based on an alert management paradigm.

8.1.5. RTRA Digiteo

Participants: Golnaz Karbaschi, 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 "Resource management in delay tolerant networks" was selected, which is financing a 1-year Post-Doc fellowship of Golnaz Karbaschir in ASAP Saclay. The project targets the use of network coding to improve the use the network availabilities, like throughput. Khaldoun Al Agha and Steven Martin from the LRI/University of Paris-Sud are also part of the project.

8.1.6. ANR Project USS-SimGrid

Participant: Fabrice Le Fessant.

USS-SimGrid has been accepted in the ANR call “Embedded Systems and Big Infrastructures”. It is composed of seven academic research labs, aiming at improving the SimGrid grid simulator. The main goal is to increase the scalability of the simulator to be able to simulate bigger and more asynchronous systems, such as peer-to-peer systems. The ASAP team is responsible for testing the simulator by using it to simulate collaborative backup and peer-to-peer video-on-demand protocols. The project is a 3-year project starting at the end of 2008.

8.1.7. ICOM Project of the PICOM

Participant: Aline Carneiro Viana.

ICOM (“Infrastructure pour le COMmerce du futur”) is a project of the “Pôles de Compétitivé Industries du commerce (PICOM)”. This project is composed by three INRIA research teams (POPS, ASAM, and ASAP) and by some big industrial and commercial partners (i.e. Declathon, La Redoute, Orange, La Poste, Auchan, GS1, Atos Origin). This project targets the study, the exploitation, and the experimentation of a large-scale ambient platform that will (1) make simple the fast deployment of new ubiquitous applications of infrastructure management and (2) cope with heterogeneous technologies (RFID, NFC, Zigbee, etc). This project has officially started in April 2008. A 2-year research engineer will be selected to work on the project.

8.1.8. Project RIAM-Solipsis

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

This project aims at designing a virtual world system such as *Second Life* in a fully distributed way. It deals with social usages, 3D-modelisation of objects in the virtual world, and the peer-to-peer infrastructure. The role of ASAP 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. 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.

8.1.9. 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. The role of the ASAP project-team is to provide the peer-to-peer infrastructure to distribute large multimedia files in an efficient way, in particular using epidemic protocols to achieve efficient and relevant clustering. The following partners within the *Pôle de Competitivité Images & Réseaux* are involved : Thomson R&D, Thomson Broadcast & Multimedia, Mitsubishi Electric ITE/TCL, Devoteam, France Télécom, ENST Bretagne, Marsoin, IPdiva, TMG and eOdus.

8.2. International grants

8.2.1. GOSSPLE ERC Starting Grant

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

Anne-Marie Kermarrec is the principal investigator of the GOSSPLE ERC tarting Grant (Sept.2008-Sept 2013). Gossple aims at providing a radically new approach to navigating the digital information universe. This project has been granted a 1.250.000 euros budget for 5 years.

Gossple aims at radically changing the navigation on the Internet by placing users affinities and preferences at the heart of the search process. Complementing traditional search engines, Gossple will turn search requests into live data to seek the information where it ultimately is: at the user. Gossple precisely aims at providing a fully decentralized system, auto-organizing, able to discover, capture and leverage the affinities between users and data.

Complementing Google-like search engines, Gossple will turn the request into a dynamic object navigating the network using epidemic protocols to find matching users/data. relevant clusters of users and data. At the heart of this procedure lies dynamic overlays based on users affinities, preferences and recommendations.

This goes far beyond discovering indexed data. In a nutshell this can be summarized in four words: decentralised nature, scalability, adaptation to dynamics and reactivity. Providing distributed algorithms that let global properties emerge from individual decisions, connecting billions of users and/or objects and navigating efficiently in such extremely large digital universe, capturing and leveraging affinities, habits, recommendations, coping with arbitrary behaviours, providing efficient protocols to match the offer and the supply are as many difficult issues to be solved. In short, the expected outcome of Gossple are theoretical advances in large-scale distributed computing, practical advances in large-scale distributed systems, an innovative software to create dynamic autonomous networked systems and the ground for new kind of distributed collaborative applications. Epidemic protocols and peer to peer overlay networks are at the core of Gossple.

8.2.2. ReSIST European project

Participants: Marin Bertier, François Bonnet, Achour Mostefaoui, Michel Raynal.

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.

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

Participants: François Bonnet, Kevin Huguenin, Anne-Marie Kermarrec, Fabrice Le Fessant, Cigdem Sengul, Aline Carneiro Viana.

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.

Following the major event for Epi-Nets that we organized in Leiden in December 2006 and 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, one of the highlight in 2008-2009 is the Special Issue of Computer Network Elsevier Journal, to appear in 2009,

Two main scientific activities were carried out in 2008. The first one is related to epidemics for video on demand ((Kevin Huguenin (ASAP), Anne-Marie Kermarrec (ASAP), Maarten van Steen (VU), Vivek Rai (VU)). The joint work focused on incentive-based video-on-demand (VoD). The second is related to epidemics for efficient streaming: PULP (Anne-Marie Kermarrec (ASAP), Etienne Riviere (ASAP, now Neuchatel), Spyros Voulgaris (ETH, now VU). In collaboration with Pascal felber, University of Neuchatel.

This year, Kevin Huguenin visited Vrije Universiteit for a month in June 2008. Maarten van Steen visited INRIA in September 2008. Fabrice le Fessant, Aline Viana and Cigdem Sengul visited Vrije Universiteit for a week in October 2008.

8.3. Visits (2008)

Artur Ziviani, LNCC, Brazil, September 2008 - January 2009, Digiteo-Labs Grant

Maarten van Steen, VU University, Amsterdam (NL), 9-10 Septembre 2008

Mark Jelasity, University of Szeged, Hungary, 9-11 September 2008

Shen Lin, University of Lancaster, UK, June 2008

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 is the Guest Editor with Maarten van Steen of the Computer Networks Elsevier Journal, Special Issue on Gossip-based networking, Spring 2009.

She served in the program committees for the following conferences:

ICDE'08: *IEEE 24th Int'l Conference on Data Engineering (ICDE 200))*, Distributed, Parallel, and Peer-to-Peer Databases Track, Cancoun, 2008.

IPDPS'08: *IEEE Int'l Conference on Parallel and Distributed Systems*, Miami, Florida, Sep 2008.

EuroSys'08: *European Conference on Computer Systems*, Glasgow, Scottland, Apr 2008.

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

SASO'08: *IEEE Int'l Conference on Self-Adaptive and Self-Organizing Systems*, Venice, Italy, Oc 2008.

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

DEBS'08: *2nd Int'l Conference on Distributed Event-Based Systems*, Rome, Italy, Jul 2008.

OPODIS'08: *Int'l Conference on Distributed Systems*, Louxor Egypt, Dec 2008.

P2P'08: *IEEE Conference on Peer-to-Peer systems*, Aachen, Germany, Sep 2008.

Algotel08 *Rencontres Francophones sur les Aspects Algorithmiques de Télécommunications*, Saint-Malo, France, May 2008.

CARI08 *Colloque Africain sur la Recherche en Informatique et en Mathématiques Appliquées*, Rabat, Morocco, Oct 2008.

CoNext08 *ACM Conext*, Madrid, Spain, Dec 2008.

ICPP08 *Int'l Conference on Parallel Processing*, Portland, Oregon, USA, Sep 2008.

WRAITS08 *Workshop on Recent Advances on Intrusion-Tolerant Systems*, Glasgow, Scotland, March 2008.

ICDCS'09: *Int'l Conference on Distributed Computing Systems* , Operating Systems and Middleware Track, Montreal, Canada, Jun 2009.

IPDPS'09: *IEEE Int'l Conference on Parallel and Distributed Systems*, Rome, Italy, May 2009.

Infocom'09: *IEEE Conference on Computer Communications and Networking*, Rio De Janeiro, Brazil, Apr 2009.

PODC'09: *ACM Conference on Principles on Distributed Computings*, Aug 2009.

Aline C. Viana served in the program committees for the following conferences/workshops :

WGRS'08: *13th Workshop on Management and Operation of Networks and Services* (WGRS 2008), in conjunction with the SBRC 2008, Rio de Janeiro, Brazil, May 2008.

CONEXT'08: *ACM CoNEXT Student Workshop*, Madrid, Spain, Dec 2008.

IMAGINE'08: *ICALP Workshop on Mobility, Algorithms and Graph theory in dynamic Networks*, Reykjavik, Iceland, Jul 2008.

PIMRC'08: *IEEE Int'l Symposium on Personal, Indoor and Mobile Radio Communications*, Cannes, France, Sep 2008.

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

P2P'08: *IEEE Conference on Peer-to-Peer Systems* (P2P 2008), Aachen, Sep 2008.

DaMaP'08: *EDBT Workshop on Data Management in P2P Systems*, Nantes, Jun 2008.

HotP2P'09: *IPDPS Workshop on Hot Topics in P2P Systems*, Roma, May 2009.

DaMaP'09: *EDBT Workshop on Data Management in P2P Systems*, Saint Petersbourg, Mar 2009

Michel Raynal served in the program committees of the following conferences:

EDCC'08 *7th European Dependable Computing Conference* . Proceedings published by Springer-Verlag LNCS.

ICDCS'08 **Chair** of the track *Theoretical Foundations* of the *28th IEEE Int. conf. on Distributed Computing Systems , Beijing (China), june 2008*. Proceedings published by IEEE Computer Society Press.

ICA3PP'08 *8th Int'l Conference on Algorithms And Architectures for Parallel Processing*. Proceedings published by Springer-Verlag LNCS. 9-11 June 2008, Cyprus.

NCA'08 *7th IEEE Int'l Symposium on Network Computing and Applications*. Proceedings published by IEEE Computer Society Press, Boston (MA), July 2008.

DISC'08 *22th Int'l Symposium on Distributed Computing*. Proceedings published by Springer-Verlag LNCS, October 2008.

HPCC'08 *10th IEEE Int'l Conference on High Performance Computing and Communications*. Proceedings published by IEEE Computer Society Press, September 2008.

He served also in the editorial board of the following journals:

JPDC Journal of Distributed and Parallel Computing

IEEE TPDS IEEE Transactions on Parallela nd Distriubted Systems

JCIS Journal of Computer and Information Sciences.

IJCSSE Int'l Journal of Computer Systems Science and Engineering

He served also in the steering committees of the following conferences:

PODC'08 "ACM Conference on the Principles of Disitributed Computing"

SIROCCO'08 "Int'l Colloquium on Structural Information and Communication Complexity"

ICDCN'08 "Int'l Conference on Distributed Computing and Networking"

ICDCS'08 "IEEE Int'l Conference on Distributed Computing Systems"

TCDP Member of the Advisory Board of the Technical Committee on Distributed Processing (TCDP) of the IEEE Computer Society. European representative.

A. Mostefaoui served in the program committees for the following conferences:

SIROCCO'08 *15th International Colloquium on Structural Information and Communication Complexity*, June 2008, Villars sur Ollon.

SSS'08 *10th International Symposium on Stabilization, Safety and Security of Distributed Systems*, November 2008, Detroit

EACP'08 *International Workshop for Communities of Practice*

ISPS'09 *9th International Symposium on Programming and Systems*, May 2009, Algiers.

He was also the co-Chair of the Workshop on Reliability in Decentralized Distributed Systems (OTM Workshops), November 2008, Monterrey.

9.1.3. Evaluation committees, consulting

Fabrice Le Fessant served as a reviewer for two projects in the ANR Arpège call.

He is a member of the Project Committee of the working group on Free Software in the Pôle de Compétitivité System@tic.

A.-M. Kermarrec served as a reviewer of the HAGGLE IP EC-funded project.

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

Yann Busnel served as a reviewer for one project in the Aquitaine Region research call.

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.

Achour Mostefaoui is heading the Master SSI (Security of Computing Systems) of the University of Rennes. He is also responsible of a Master's course (University of Bougie, Algeria) entitled "Distributed Algorithms and Systems".

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)

Aline C. Viana is a teacher in the Master's courses (M2R) at the University of Paris-Sud.

Fabrice le Fessant is a half-time associate professor at Ecole Polytechnique.

Yann Busnel acted as Teaching Assistant (*moniteur*) both in UNIVERSITY RENNES 1 and ENS CACHAN.

In addition, among Ph.D students, **Antoine Boutet, Nicolas Le Scouarnec** and **Vincent Leroy** are Teaching Assistants (*moniteurs*) at INSA.

9.3. Conferences, seminars, and invitations

9.3.1. Invited Talks

A.-M. Kermarrec, D. Frey and **Fabrice Le Fessant** have been invited speakers at IRISAtech.

Fabrice Le Fessant has been an invited speaker at Forum Atena.

A.-M. Kermarrec and **Michel Raynal** have been invited speakers at the INRIA Winter School on Hot Topics in Distributed Computing at La Plagne in March 2008.

A.-M. Kermarrec has been invited to give a tutorial at the Int'l Conference on Event-Based systems (DEBS) on large-scale publish-subscribe systems, at Rome in July 2008.

She has also been invited to give a talk at the METIS Summer School on "Peer to peer Networking" in Morocco in May 2008.

Yann Busnel has been invited to spends some weeks at the Middleware Laboratory, Department of Computer Science, University of Rome "La Sapienza", to work in collaboration with Roberto Baldoni, in November 2007, December 2007, June 2008 and from September 2008 to December 2008.

He also obtained a Student Scholarship to assist at the 21st ACM Symposium on Operating Systems Principles (SOSP'07), in Stevenson, WA, USA, in October 2007

9.3.2. Seminars

A.-M. Kermarrec has been invited to give a talk on "Distributed slicing"at the LADIS workshop, New York, September 2008.

Fabrice Le Fessant, Cigdem Sengul and **Aline C. Viana** have been invited to give a seminar at the Department of Computer Science, Vrije Universiteit Amsterdam, in October 2008.

Yann Busnel has been invited to give two seminars at the Department of Computer Science, University of Rome "La Sapienza", respectively in November 2007 and in December 2008.

9.4. Administrative responsibilities

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

She was a member of the 2008 INRIA Selection Committees for Senior Researcher permanent positions (DR2), and for Junior Researcher permanent positions (CR2) (INRIA RENNES - BRETAGNE ATLANTIQUE and INRIAGRENOBLE - RHÔNE-ALPES).

Aline C. Viana is a member of the "Commission développement technologique" of the INRIA.

Fabrice Le Fessant is in the Free Software Project Committee of Pôle de Compétitivité System@tic.

Yann Busnel is the national representative of the CIES Grand Ouest (French University Professor Education Agency) and an Active Member of CJC (The French Confederation of Young Researcher)

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

Year Publications

Articles in International Peer-Reviewed Journal

[8] R. BALDONI, F. BONNET, A. MILANI, M. RAYNAL. *Anonymous graph exploration without collision by mobile robots*, in "Information Processing Letters", 2008, http://hal.inria.fr/inria-00333179/en/.

[9] J. RAHME, A. CARNEIRO VIANA, K. AL AGHA. *Looking for network functionalities' extension by avoiding energy-compromised hotspots in wireless sensor networks*, in "Springer Annals of Telecommunication Journal", vol. 63, 2008, http://hal.inria.fr/inria-00334342/en/.

[10] C. SENGUL, M. MILLER, I. GUPTA. *Adaptive Probability-Based Broadcast Forwarding in Energy-Saving Sensor Networks*, in "Transactions on Sensor Networks,", 2008, http://hal.inria.fr/inria-00333703/en/.

Articles in Non Peer-Reviewed Journal

[11] G. TRÉDAN, A. CARNEIRO VIANA. *Systemes repartis de grande taille: de l'anarchie a l'auto-structuration*, in "Reseaux et Telecoms - Lettre Bimestrielle", 2008, http://hal.inria.fr/inria-00334350/en/.

Invited Conferences

[12] D. FREY, R. GUERRAOUI, A.-M. KERMARREC, M. MONOD, V. QUEMA. *Heterogeneous Gossiping*, in "Large-Scale Distributed Systems and Middleware, États-Unis d'Amérique IBM TJ Watson Research Lab in Yorktown, New York.", 2008, http://hal.inria.fr/inria-00337056/en/.

International Peer-Reviewed Conference/Proceedings

[13] F. BONNET, M. RAYNAL. *Conditions for Set Agreement with an Application to Synchronous Systems*, in "ICDCS, Chine Beijing", 2008, http://hal.inria.fr/inria-00333162/en/.

[14] Y. BUSNEL, M. BERTIER, A.-M. KERMARREC. *Solist or How To Look For a Needle in a Haystack? A Lightweight Multi-Overlay Structure for Wireless Sensor Networks*, in "IEEE WiMob '08, France Avignon", 2008, http://hal.inria.fr/inria-00338128/en/.

[15] Y. BUSNEL, L. QUERZONI, R. BALDONI, M. BERTIER, A.-M. KERMARREC. *On the Deterministic Tracking of Moving Objects with a Binary Sensor Network*, in "ACM/IEEE DCOSS '08, Grèce Ile Santorin", 2008, http://hal.inria.fr/inria-00338129/en/.

[16] A. CARNEIRO VIANA, M. DIAS DE AMORIM. *Sensing and acting with predefined trajectories*, in "ACM Workshop on Heterogeneous Sensor and Actor Networks (ACM HeterSanet), in conjunction with ACM Mobihoc, Chine Hong Kong", 2008, http://hal.inria.fr/inria-00334357/en/.

[17] D. FREY, A. MURPHY. *Failure-Tolerant Overlay Trees for Large-Scale Dynamic Networks*, in "8th International Conference on Peer-to-Peer Computing 2008 (P2P'08), Allemagne Aachen", 2008, http://hal.inria.fr/inria-00337054/en/.

[18] D. FREY, J. ROYAN, R. PIEGAY, A.-M. KERMARREC, E. ANCEAUME, F. LE FESSANT. *Solipsis: A Decentralized Architecture for Virtual Environments*, in "1st International Workshop on Massively Multiuser Virtual Environments, États-Unis d'Amérique Reno, NV", 2008, http://hal.inria.fr/inria-00337057/en/.

[19] A.-M. KERMARREC, A. MOSTEFAOUI, M. RAYNAL, G. TRÉDAN, A. CARNEIRO VIANA. *(BA) From Anarchy to Geometric Structuring: the Power of Virtual Coordinates*, in "Proceedings of ACM Sigact-Sigops Symposium on Principles of Distributed Computing (PODC), Canada Toronto", 2008, http://hal.inria.fr/inria-00334356/en/.

[20] A.-M. KERMARREC, A. MOSTEFAOUI, M. RAYNAL, G. TRÉDAN, A. CARNEIRO VIANA. *Large-scale networked systems: from anarchy to geometric self-structuring*, in "10th International Conference on Distributed Computing and Networking (ICDCN), Inde Hyderabad", 2009, http://hal.inria.fr/inria-00334355/en/.

[21] E. LE MERRER, A.-M. KERMARREC, V. GRAMOLI. *Distributed churn measurement in arbitrary networks*, in "PODC '08: The twenty-seventh ACM symposium on Principles of distributed computing, Canada Toronto, Canada", ACM, 2008, p. 431–431, http://hal.inria.fr/inria-00329932/en/.

[22] E. LE MERRER, A.-M. KERMARREC, V. GRAMOLI. *Méthode décentralisée de mesure de dynamisme d'un réseau logique*, in "6eme Conference Francaise en Systemes d'Exploitation (CFSE'08), Suisse Fribourg", ACM SIGOPS de France, 2008, http://hal.inria.fr/inria-00329946/en/.

[23] E. LE MERRER, A.-M. KERMARREC, V. GRAMOLI. *SONDe, a Self-Organizing Object Deployment Algorithm in Large-Scale Dynamic Systems*, in "EDCC 2008. Seventh European Dependable Computing Conference,", IEEE, 2008, p. 157-166, http://hal.inria.fr/inria-00329928/en/.

[24] E. LE MERRER, A.-M. KERMARREC, G. TRÉDAN, B. SERICOLA. *Evaluating the Quality of a Network Topology through Random Walks*, in "DISC 2008: 22nd International Symposium on Distributed Computing, France Arcachon", Springer, 2008, p. 509–511, http://hal.inria.fr/inria-00329941/en/.

[25] A. MOSTEFAOUI, M. RAYNAL, C. TRAVERS. *Narrowing Power vs. Efficiency in Synchronous Set Agreement*, in "9th International Conference on Distributed Computing and Networking LNCS, Inde Kolkata", M. C. SHRISHA RAO, P. JAYANTI, C. S. R. MURTHY, S. K. SAHA (editors), vol. 4904, Springer Verlag, 2008, p. 99-111, http://hal.inria.fr/inria-00338093/en/.

[26] C. SENGUL, M. BAKHT, A. HARRIS, T. ABDELZAHER, R. KRAVETS. *Improving Energy Conservation Using Bulk Transmission over High-Power Radios in Sensor Networks*, in "International Conference on Distributed Computing Systems, Chine pekin", 2008, http://hal.inria.fr/inria-00333700/en/.

[27] G. TAN, S. A. JARVIS, A.-M. KERMARREC. *Connectivity-Guaranteed and Obstacle-Adaptive Deployment Schemes for Mobile Sensor Networks*, in "International Conference on Distributed Computing Systems (ICDCS 2008), Chine Beijing", 2008, http://hal.inria.fr/inria-00338592/en/.

National Peer-Reviewed Conference/Proceedings

[28] Y. BUSNEL. *SOLIST : Structure multi-couche pair-à-pair à faible consommation pour les réseaux de capteurs sans-fil*, in "Conférence Francophone sur les Systèmes d'Exploitation - 6ème édition, Suisse Fribourg", 2008, http://hal.inria.fr/inria-00338130/en/.

[29] Y. BUSNEL, L. QUERZONI, R. BALDONI, M. BERTIER, A.-M. KERMARREC. *Suivi et identification de trajectoires sur un réseau de capteurs binaires*, in "AlgoTel '08, France Saint-Malo", 2008, http://hal.inria.fr/inria-00338131/en/.

[30] V. LEROY, M. BERTIER, A.-M. KERMARREC. *Vers un desktop-grid pair-à-pair*, in "CFSE '6, Suisse Fribourg", 2008, http://hal.inria.fr/inria-00331824/en/.

[31] G. TRÉDAN, A. CARNEIRO VIANA. *VINCOS – Systémes répartis de grande taille: de l'anarchie à l'auto-structuration*, in "Colloque Francophone sur l'Inglénierie des Protocoles (CFIP), France Les Arcs", 2008, http://hal.inria.fr/inria-00334351/en/.

Scientific Books (or Scientific Book chapters)

[32] A. CARNEIRO VIANA, M. BERTIER. *Data aggregation as an essential paradigm in Wireless Sensor Networks*, in "Ad Hoc Networks: New Research Ad Hoc Networks: New Research", J. N. TURNER, C. S. BOYER (editors), Nova Science Publisher, Inc., 2008, p. 45-60, http://hal.inria.fr/inria-00334346/en/.

[33] M. DIAS DE AMORIM, F. BENBADIS, M. SICHITIU, A. CARNEIRO VIANA, Y. VINIOTIS. *Routing in Wireless Self-Organizing Networks*, in "Adaptation and Cross Layer Design in Wireless Networks", M. IBNKAHLA (editor), vol. 2, Taylor and Francis Group, LLC, 2008, p. 325-353, http://hal.inria.fr/inria-00334344/en/.

Research Reports

[34] R. BALDONI, F. BONNET, A. MILANI, M. RAYNAL. *Anonymous Graph Exploration without Collision by Mobile Robots*, PI 1886, Rapport de recherche, 2008, http://hal.inria.fr/inria-00250153/en/.

[35] R. BALDONI, F. BONNET, A. MILANI, M. RAYNAL. *On the Solvability of Anonymous Partial Grids Exploration by Mobile Robots*, PI 1892, Rapport de recherche, 2008, http://hal.inria.fr/inria-00277344/en/.

[36] Y. BUSNEL, M. BERTIER, A.-M. KERMARREC. *Bridging the Gap between Population and Gossip-based Protocols*, RR-6720, Rapport de recherche, 2008, http://hal.inria.fr/inria-00338098/en/.

[37] Y. BUSNEL, M. BERTIER, A.-M. KERMARREC. *On the Impact of the Mobility on Convergence Speed of Population Protocols*, RR-6580, Rapport de recherche, 2008, http://hal.inria.fr/inria-00296584/en/.

[38] D. IMBS, M. RAYNAL. *A Lock-based Protocol for Software Transactional Memory*, PI 1893, Rapport de recherche, 2008, http://hal.inria.fr/inria-00277378/en/.

[39] D. IMBS, M. RAYNAL. *Provable STM Properties: Leveraging Clock and Locks to Favor Commit and Early Abort*, PI 1894, Rapport de recherche, 2008, http://hal.inria.fr/inria-00281550/en/.

[40] A.-M. KERMARREC, E. LE MERRER, B. SERICOLA, G. TRÉDAN. *Evaluating topology quality through random walks*, RR-6534, Rapport de recherche, 2008, http://hal.inria.fr/inria-00279074/en/.

[41] F. LE FESSANT, C. SENGUL, A.-M. KERMARREC. *Pace-Maker: Tracking peer availability in large networks*, RR-6594, Rapport de recherche, 2008, http://hal.inria.fr/inria-00305620/en/.

[42] Y. MOSES, M. RAYNAL. *No Double Discount: Condition-based Simultaneity Yields Limited Gain*, PI 1898, Rapport de recherche, 2008, http://hal.inria.fr/inria-00293649/en/.

[43] Y. MOSES, M. RAYNAL. *Revisiting Simultaneous Consensus with Crash Failures*, PI 1885, Rapport de recherche, 2008, http://hal.inria.fr/inria-00260643/en/.

[44] A. MOSTEFAOUI, S. RAJSBAUM, M. RAYNAL, C. TRAVERS. *The Combined Power of Conditions and Information on Failures to Solve Asynchronous Set Agreement*, PI 1897, Rapport de recherche, 2008, http://hal.inria.fr/inria-00294706/en/.

References in notes

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

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

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

[48] A. DEMERS, D. GREENE, C. HAUSER, W. IRISH, J. LARSON. *Epidemic algorithms for replicated database maintenance*, in "Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC'87)", August 1987.

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

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

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

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

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

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

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

[56] 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.