INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Team Regal

Resource management in large scale distributed systems

Rocquencourt

THEME 1A

Activity Report

2003

Table of contents

1. Team

Regal is a joint team between INRIA, CNRS and Paris 6 University, through the "Laboratoire d'Informatique de Paris 6", LIP6 *(UMR 7606).*

Head of project-team

Pierre Sens [Professor University of Paris 6]

Vice-head of project-team

Mesaac Makpangou [Research associate (CR) INRIA]

Administrative assistant

Noelle Bourgeois [ITA]

Staff member LIP6

Luciana Arantes [Associate professor University of Paris 6]
Bertil Folliot [Professor Univeristy of Paris 6]

Engineers

Simon Patarin [from June to October (INRIA)]
Ian Puimarta [LIP6]

Ph. D. student

Carine Baillarguet [INRIA]
Marin Bertier [University of Paris 6]
Jean-Michel Busca [University of Paris 6]
Ikram Chabbouh [University of Tunis]
Corina Ferdean [INRIA]
Nicolas Gibelin [INRIA]
Ahmed Jebali [INRIA (until November)]
Assia Hachichi [University of Paris 6]
Christian Khoury [University de Paris 6]
Olivier Marin [University de Paris 6, University of Le Havre (until November)]
Ahmed Mokhtar [INRIA (until September)]
Le Chau Nguyen Thi [INRIA]
Simon Patarin [INRIA (until may)]
Fabio Piconni [University of Paris 6]
Gaël Thomas [University of Paris 6]

2. Overall Objectives

Regal is a joint research team between the LIP6 and INRIA-Rocquencourt. Regal focuses on management of large scale distributed systems (especially P2P architectures). A significant axis of the project work is devoted to large scale replication for the applications which have strong constraints in terms of dynamicity (multi-agents applications and servers of resources on the Web).

3. Scientific Foundations

Key words: *distributed system*, *replication*, *consistency*, *fault tolerance*, *dynamic adaptation*, *large scale environments*, *Peer-to-peer*, *Grid computing*.

We target highly distributed environments whom configurations include large number of processors and long distances between hosts. Such environments which are highly dynamic, heterogeneous, and rule out the simple possibility of a centralized and/or instantaneous management of a global knowledge.

Regal proposes the use of of reactive replication to tolerate faults and to reduce the access time to information, while adapting dynamically to the environmental constraints and the evolution of application behavior. Regal concentrates on the deployment of applications (code and data) adapted to highly distributed environments.

We concentrate on the following research themes:

Data management: the goal is to be able to deploy and locate effectively data while maintening the required level of consistency between data replicas.

System monitoring and failure detection: we envisage a service providing the follow-up of distributed information. Here, the first difficulty is the management of a potentially enormous flow of information which leads to the design of dynamic filtering techniques. The second difficulty is the asynchronous aspect of the underlying network which introduces a strong uncertainty on the collected information.

Adaptive replication: we design parameterizable techniques of replication aiming to tolerate the faults and to reduce information access times. We focus on the runtime adaptation of the replication scheme by (1) automatically adjusting the internal parameters of the strategies and (2) by choosing the replication protocol more adapted to the current context.

The dynamic adaptation of application execution support: the adaptation is declined here to the level of the execution support (in either of the high level strategies). We thus study the problem of dynamic configuration at runtime of the low support layers.

4. Application Domains

Key words: *Internet services*, *data storage*, *multi-agent systems*.

We target two kinds of large scale environments : computional grids and peer-to-peer (P2P) systems. While the two environments have the same final objective to share large sets of resources, they initially emerged from different communities with different context assumptions and hence they have been designed differently. Grids provide support for a large number of services needed by scientific communities. They usually target thousands of hosts and hundred of users. Peer-to-peer environments address millions of hosts with hundreds of thousands of simultaneous users but they offer limited and specialized functionalities (file sharing, parallel computation).

In peer-to-peer configuration we focus on the following applications :

- Internet services such as web caches or content distribution network (CDN) which aim to reduce the access time to data shared by many users,
- Data storage of mutable data. Data storage is a classical peer-to-peer application where users can share documents (audio and video) across Internet. A challenge of the next generation of data sharing system is to provide update facility in order to develop large cooperative application.

In Grid configuration we address the resource management for two kinds of applications :

- Multi-agent applications which model complex cooperative behaviors.
- Application Service Provider (ASP) environments in cooperation with the DIET project of the GRAAL team.

5. Software

5.1. DARX (Dynamic Agent Replication eXtension)

Participants: Pierre Sens [correspondent], Marin Bertier, Olivier Marin.

DARX is a framework for building applications that provides adaptive fault tolerance. It relies on the fact that multi-agent platforms constitute a very strong basis for decentralized software that is both flexible and scalable, and makes the assumption that the relative importance of each agent varies during the execution. DARX groups solutions which facilitate the creation of multi-agent applications in a large-scale context. Its most important feature is adaptive replication: replication strategies are applied on a per-agent basis according to transient environment characteristics such as the criticity of the agent for the computation, the network load or the mean time between failures.

DARX constitutes a solution for the automation of adaptive replication schemes, supported by a low-level architecture which addresses scalability issues. The latter is composed of several services.

- A failure detection service which maintains the dynamic lists of all the running DARX servers as well as of the valid replicas which participate to the current application, and notifies the latter of suspected failure occurrences.
- A naming and localization service generates a unique identifier for every replica in the system, and returns the addresses of all the replicas of a same group as a response to an agent localization request.
- A system observation service monitors the behavior of the underlying distributed system: it collects low-level data by means of OS-compliant probes

5.2. Pandora

Participants: Mesaac Makpangou [correspondent], Simon Patarin.

Pandora is a general purpose monitoring platform. It offers a high level of flexibility, while still achieving good performance. Each monitoring task executed by Pandora is splitted into basic and self-contained building blocks called components. These components are chained within stacks to constitute high level tasks. Stack execution consists of components exchanging messages – data structures called "packets" – from the beginning of the stack, to the end.

Pandora provides a framework dealing mainly with packet demultiplexing, timers, threads and communication management. This allows programmers to concentrate on the precise functionalities they want to implement and promotes code reuse.

Pandora has been used by many applications:

- Monitoring of HTTP traffic,
- On-line measurement of web proxy cache efficiency,
- Analysis of the MLdonkey peer-to-peer file-sharing program.

5.3. Sabbarus

Participants: Mesaac Makpangou [correspondent], Ahmed Jebali.

The Sabbarus framework support a replication scheme in which replicas need not be connected to the whole network all the time. Sabbarus is written on top of Ensemble. It performs a simple, and compact interface to write and administrate replicated application. The main functionalities of Sabbarus are :

- a library for developing clients application accessing a replicated object,
- a mechanism for defining specific access to the replicated object,
- a command line interface for controlling the framework.
- a mechanism for controlling the degree of divergence accepted by the application (according to its semantic).

6. New Results

6.1. Introduction

In 2003, we focus our research on the following areas :

- Distributed algorithm for Grid configuration.
- Data storage in peer-to-peer system.
- Dynamic adaption of virtual machines.
- Deployment of decentralized applications
- Replication strategies.

6.2. Distributed algorithms

In this theme, we study the adaptation of classical distributed algorithms to large scale configurations. Mutual exclusion is a basic block of distributed synchronization algorithms. One of the challenges in highly distributed environments (like peer-to-peer or Grid configurations) is to provide scalable synchronization taking into account the hierarchical network topology. We proposed a number of hierarchical mutual exclusion algorithms. These algorithms are extensions of the Naimi-Trehel's token algorithm, reducing the latency and the number of messages exchanged between distant hosts. These extension follow mainly three approaches: (1) an approach based on hierarchical proxy-based approach, (2) the aggregation of requests, and (3) token preemption by closer hosts.

We compared the performance of these algorithms on an emulated Grid testbed. We studied the impact of each of the extensions, showing that the combination of them can greatly improve the performance of the original algorithm.

6.3. Storage in peer-to-peer systems

In 2003, we developed two file systems for peer-to-peer configurations : POST in cooperation with Rice University and Pastis.

POST is a decentralized messaging infrastructure that supports a wide range of collaborative applications, including electronic mail, instant messaging, chat, news, shared calendars and whiteboards. POST is highly resilient, secure, scalable and does not rely on dedicated servers. Instead, POST is built upon a peer-to-peer (p2p) overlay network whom participants desktop computers. POST offers three simple and general services, such as (i) secure, single-copy message storage, (ii) meta-data based on single-writer logs, and (iii) event notification.

Pastis is a completely decentralized multi-user read-write peer-to-peer file system. In our system every file is described by a modifiable inode-like structure which contains the addresses of the immutable blocks in which the file contents are stored. All data is stored using the Past distributed hash table, which is built on top of the Pastry peer-to-peer network overlay. Authentication and integrity is assured using standard cryptographic mechanisms. Our system takes advantage of the fault tolerance and locality properties of both Past and Pastry. This allows us to optimize message routing and replica retrieval so that the performance cost caused by network latency is kept at a minimum. We have also introduced a modification to the Past DHT which allows us to further increase performance when using a relaxed but nevertheless reasonable consistency model. We have developed a prototype in order to evaluate the performance of our design. Our prototype is programmed in Java and uses the FreePastry open-source implementation of Past and Pastry. It currently provides a proprietary Java interface to applications, and allows them to choose between two degrees of consistency. Preliminary results obtained suggest that our system is approximately twice as slow as NFS.

6.4. Virtual virtual machine (VVM)

The VVM team works on flexible execution environments, based on a HAL (Hardware Abstraction Layer) and a flexible dynamic compiler, free of any predefined, imposed abstractions. This high level of dynamic flexibility allows the dynamic construction of dedicated execution environments as well as their dynamic reconfiguration. To demonstrate the benefits of this approach, we used this minimal execution environment to build the JNJVM (JNJVM is Not a JVM), a dynamically adaptable Java runtime. The inherent flexibility of the JNJVM allows us to address some of the limitations of traditional rigid Java runtimes, in particular concurrency management and concurrent programming support. Other works concern A2N (Active Active Network), a minimal active node architecture that has the capability of being dynamically specializable at runtime through capsules, and C/SPAN a self-adapting Web proxy cache (based on PANDORA) that applies administrative strategies to adapt itself and react to external events. Because C/SPAN is completely flexible, even these adaptation policies can be dynamically adapted.

Within the VVM project we continue to investigate a systematic approach for building flexible, adaptable and interoperable execution environments, to free applications from the artificial limitations on reconfiguration imposed by programming environments. The HAL part of the minimal execution environment is currently being embedded in a Linux kernel-module so that our flexible Java runtime can be used on top of a traditional operating system, as a standard JVM, while preserving a maximum of flexibility. Concerning the JNJVM, we plan to further develop our flexible Java environment towards a complete flexible Java-OS and also to integrate it in the context of component based middleware, such as the Corba Component model, to illustrate the adequacy of our approach to the needs of modern distributed applications. This work is being made within the IST COACH European project.

6.5. Deployment of decentralized applications

We propose a system and a domain specific language to facilitate the deployment of heterogeneous shared applications on heterogeneous machines. Such a situation is typical in today's Internet. The objective of the proposed language and system support is to reduce the burden on providers and clients of applications that are made available on the Internet and have certain requirements in terms of access performance availability , etc. The proposed domain specific language allows each provider to specify the application dependencies, as well as the adaptations induced by the characteristics of the hosting system. Moreover, this language allows the description of all the components of an application, including the components interfaces and hooks. The specification permits the description of the replicated applications, but the replication remains transparent to the user. The deployment system exploits this specification to build at run-time, for each application, an executive environment that offers all the services required by the applications. The proposed system also provides a mean to control the life-cycle of application releases within the hosting infrastructure.

6.6. Replication strategies

The effectiveness of replication, as a key technique to improve the clients perceived QoS when accessing large-scale distributed services, depends on decisions such as the selection of the suitable replica to assign to a client, when to create new replicas, where to place new replicas or when existing replicas become useless. Our work consists in developing a dynamic replication infrastructure aimed to improve the service delivered QoS (in terms of availability, throughput, and response time), as perceived by the clients, while consuming as little system and network resources as possible. We base our decisions for the replica management on the providers/clients requirements, aggregated in a replication contract, comprising policies for replica selection, placement, creation, migration, elimination, and consistency.

Concerning the replica selection problem, the criteria for assigning replicas to requesting clients should consider the metrics that contribute to the user-perceived QoS and some means to estimate this metrics for each replica. For example, a replicated Web service, aimed to minimize the documents transfer time, should consider link latency metric when distributing small documents and the available bandwidth for large documents (as in this case, the competing traffic becomes the primary factor limiting the document transfer time. Also, if the client requests involve intensive computing tasks on the replica server, the replica host load is an important metric to be taken into consideration. In the general case, the applications providers and the clients have different views of what a suitable replica means, according to the application characteristics, the guarantees that the application provider wishes to offer to its clients, and the clients preferences. Unfortunately, most previous works on replica selection use the same metrics for all applications and clients. This uniformity limits the efficiency of these strategies when they face the diversity of applications to be replicated and the variety of clients preferences.

Concerning replicas placement, the application providers could have different policies in terms of preferred (geographical/topological) locations where new replicas should be placed, and also in terms of statical/dynamical physical and software resources that the replica host should own. As far as we know, existing work on replica management face the same limitation of using replication criteria predefined for all applications (existing replication criteria only adapt eventually to the clients access patterns.

Our current work on replication is focused on the design and implementation of a consistency meta-model, aiming to provide fine-grained customization for replicated data objects.

8. Other Grants and Activities

8.1. National initiatives

8.1.1. Distributed algorithms and application CNRS Grant

Members: LIAFA, IRISA (Adept Team), LRI, Regal, EPFL

Objectives: This working group leaded by Hugues Fauconier and Carole Delporte (LIAFA) focuses on the application of theorical results in distributed algorithms on the large scale environments

8.1.2. ENWEG CNRS Grant

Members: LRI, CEA, ID-IMAG, INRIA Sophia, IRISA, LABRI, LIFC, LIFL, LIP, LIP6, LORIA, LRI, PRISM

Objectives: This project leaded by Franck Cappello (LRI) aims to identify issues (scientific and technical) and proposes solutions following the overall goal of building an experimental Grid platform gathering nodes geographically distributed in France.

8.1.3. Grid Data Service - ACI MD (2003-2005)

Members: IRISA (Paris Team), ENS-Lyon (LIP - Remap Team), Regal

Objectives: The goal of this project is to propose an approach where the grid computation is decoupled from data management, by building a data sharing service adapted to the constraints of scientific grid computations. The main goal of this project is to specify, design, implement and evaluate a data sharing service for mutable data and integrate it into the DIET ASP environment developed by the ReMaP team of LIP. This service will be built using the generic JuxMem platform for peer-to-peer data management (currently under development within the PARIS team of IRISA, Rennes). The platform will be used to implement and compare multiple replication and data consistency strategies defined together by the PARIS team (IRISA) and by the REGAL team of LIP6.

8.1.4. Data Grid eXplorer (2003-2005)

Members: IMAG-ID, Laria, LRI, LAAS, LORIA, LIP Ens-Lyon, LIFL, INRIA Sophie Antipolis, LIP6, IBCP, CEA, IRISA INRIA Rocquencourt

Objectives: The goal of Data Grid Explorer is to build an emulation environment to study large scale configurations. Today, it is difficult to evaluate new models for data placement and caching, network content distribution, peer-to-peer systems, etc. Options include writing simulation environments from scratch, employing detailed packet-level simulation environments such as NS, local testing within a controlled cluster setting, or deploying live code across the Internet or a Testbed. Each approach has a number of limitations. Custom simulation environments typically simplify network and failure characteristics. Packet-level simulators add more realism but limit system scalability to a few hundred of simultaneous nodes. Cluster-based deployment adds another level of realism by allowing the evaluation of real code, but unfortunately the network is highly over-provisioned and uniform in its performance characteristics. Finally, live Internet and Testbed deployments provide the most realistic evaluation environment for wide-area distributed services. Unfortunately, there are significant challenges to deploying and evaluating real code running at a significant number of Internet sites. The main benefit of emulation is the ability to reproduce experimental conditions and results.

The project is structured horizontally into transverse working groups: Infrastructure, Emulation, Network, and Applications. The Regal team is leader for the Emulation working group.

8.1.5. ACI Grid DataGraal (2002 - 2004)

Members: Regal (coordinator), projet PARIS (IRISA), LRI, Remap team (LIP Lyon), LISI (Lyon), IMAG-ID, IMAG-LSR, SMIS team, CEA, CESR, LIRMM,

Objectives: DataGraal is a discussion forum about data management, systems and applications for large scale environments. The motivation is to benefit from the experiences of different communities: system, database, and scientific applications.

The objective is to understand the link between grid computing and distributed information systems.

We investigate the impact of the large scale environment, which represents a challenge, at the same time for calculation on grid and the distributed information systems. We aim: (1) to specify the useful concepts and the common approaches on mobility, availability and distribution of huge amount of data, (2) to make emerge some federative projects from these communities.

8.2. European initiatives

8.2.1. IST Coach (2002-2004)

Members: T-Systems, Humboldt Univeristaet zu Berlin, Intracom, Lucent, Thales, CNRS, ObjectSecurity, LIP6, LIFL, FOKUS

Objectives: COACH is concerned with the rapid and cost effective development of large scale and mission critical distributed applications by using software components. These components are based on the OMG's CORBA Component Model (CCM), which is a specification for creating distributed, server-side scalable, component-based, language-neutral, transactional, multi-user, and secure applications. The key objective is to build a component framework that is well integrated with state of the art software engineering techniques like the OMG's Model Driven Architecture. The framework can rapidly transform models, architecture and design level components, as well as policies to execution level and deploy them efficiently and securely on distributed hardware platforms. This allows the developer to concentrate on the business logic instead of reinventing technical infrastructure and to reuse existing components, thus increasing software quality and greatly reducing development costs and time to market.

9. Dissemination

9.1. Program committees and responsibilities

Luciana Arantes was member of the program committee of:

- *The 2003 International Multiconference in Computer Science and Computer Engineering*. June 2003, Las Vegas, USA.

Bertil Folliot was member of the program committee of:

- ISPDC'2003 - *International Symposium on Parallel and Distributed Computing*, Iasi, 2002, Roumanie.
- *The 2003 International Multiconference in Computer Science and Computer Engineering*. Juin 2003, Las Vegas, USA.

Mesaac Makpangou was member of the program committee of:

- Medianet'2003 - *International Conference on Intelligent Access of Multimedia Documents on Internet* , June 2002, Sousse, Tunisie.

Pierre Sens was member of the program committee of:

- ISORC'2003 - *5th IEEE International Symposium on Object-Oriented Real-Time Distributed Computing*. Mai 2003, Hakodate, Hokkaido, Japon.
- CFSE'2003 - *Conférence française sur les systèmes d'Exploitation*. October 2003, Nice, France.

P. Sens is in charge of cooperation between University of Paris 6 and the Oradea University (Roumania). P. Sens is vice-chair of the French Chapter of ACM SIGOPS

9.2. PhD reviews

Bertil Folliot was PhD rewiever of :

- Luciano Barreto, "Conception aisée et robuste d'ordonnanceurs de processus au moyen d'un langage dédié, Université de Rennes I, June 2003 (advisor: G. Muller)
- Alexandre Denis, "Contribution à la conception d'une plate-forme haute performance d'intégration d'exécutifs communicants pour la programmation des grilles de calcul", IRISA, Université de Rennes 1, December 2003 (advisor: T. Priol, C. Perrez)
- Christophe Rippert, "Protection dans les architectures de systèmes flexibles", Université Joseph Fourier, October 2003 (advisor: Sacha Krakowiak)

Pierre Sens was PhD rewiever of :

- Phuong Quynh Duong, "La tolérance aux fautes adaptable pour les systèmes à composants", IMAG-LSR Lab., Institut National Polytechnique de Grenoble, December 2003 (advisor: C. Collet)
- Pierre Lombard, "NFSp: Une solution de stockage distribué pour architectures grande échelle", IMAG-ID Lab., Institut National Polytechnique de Grenoble, December 2003 (advisor: B. Plateau, Y. Denneulin)

9.3. Teaching

- Luciana Arantes :

  – Principles of operating systems in Licence d'Informatique, Université Paris 6
  – Operating systems kernel in Maîtrise d'Informatique, Université Paris 6
  – Responsible for projects in operating system, Maîtrise d'Informatique, Université Paris 6

- Bertil Folliot :

  – Principles of operating systems in Licence d'Informatique, Université Paris 6
  – Distributed algorithms and systems in Magistère d'Informatique, Université Paris 6
  – Distributed programming, DESS DLS, Université Paris 6
  – Distributed systems and client/serveur in Maîtrise d'Informatique, DEA Systèmes Répartis, Université Paris 6
  – Responsible for the DEA Systèmes Informatique Répartis, Université Paris 6, CNAM, ENST
  – Projects in distributed programming, Maîtrise d'Informatique, Université Paris 6

- Mesaac Makpangou

  – Distributed systems, DESS IRS, Université de Versailles Saint-Quentin en Yvelines
  – Performance of distributed systems, DEA MISI, Université Versailles Saint-Quentin
  – Large Scale Data sharing, DEA IFA, Université de Marne-La-Vallee
  – Distributed systems and applications, Institut Superieur de Technologies et de management (ISTM)
  – Systems and networks, Master, Pôle Universitaire Leonard de Vinci

- Pierre Sens :

  – Principles of operating systems in Licence d'Informatique, Université Paris 6
  – Operating systems kernel in Maîtrise d'Informatique, Université Paris 6
  – Responsible for Operating System and Network part of the Maîtrise d'Informatique, Université Paris 6
  – Animation of the working group on advances in distributed system, DEA SIR, Université Paris 6

10. Bibliography

Major publications by the team in recent years

[1] C. AMZA, A. L. COX, S. DWARKADAS, P. KELEHER, H. LU, R. RAJAMONY, W. ZWAENEPOEL. *TreadMarks: Shared Memory Computing on networks of Workstations.* in « IEEE Computer », number 2, volume 29, feb, 1996, pages 18-28.

[2] B. BERSHAD, S. SAVAGE, P. PARDYAK, E. SIRER, M. FIUCZYNSKI, D. BECKER, C. CHAMBERS, S. EGGERS. *Extensibility, safety and performance in the SPIN operating system.* in « ACM Operating Systems Review », number 5, volume 29, 1995, pages 267-284.

[3] K. P. BIRMAN. *Replication and Fault-Tolerance in the ISIS System.* in « ACM Operating Systems Review », number 5, volume 19, 1985, pages 79-86, Proc. 10th ACM Symp. on Operating System Principles, Orcas Island, WA, USA, December 1985.

[4] T. D. CHANDRA, S. TOUEG. *Unreliable failure detectors for reliable distributed systems.* in « Journal of the ACM », number 2, volume 43, mar, 1996, pages 225-267.

[5] D. R. ENGLER, M. F. KAASHOEK, J. JAMES O'TOOLE. *Exokernel: An Operating System Architecture for Application-Level Resource Management.* in « Proceedings of the 15th Symposium on Operating Systems Principles (15th SOSP'95), Operating Systems Review », ACM SIGOPS, pages 251-266, Copper Mountain, CO, dec, 1995, Published as Proceedings of the 15th Symposium on Operating Systems Principles (15th SOSP'95), Operating Systems Review, volume 29, number 5.

[6] M. J. FISCHER, N. A. LYNCH, M. S. PATERSON. *Impossibility of distributed consensus with one faulty process.* in « Journal of the ACM », number 2, volume 32, apr, 1985, pages 374-382.

[7] R. GUERRAOUI, B. GARBINATO, K. R. MAZOUNI. *Garf: A Tool for Programming Reliable Distributed Applications.* in « IEEE Concurrency », number 4, volume 5, oct/dec, 1997, pages 32-39.

[8] R. GUERRAOUI. *Indulgent Algorithms.* in « Proceedings of the 19th Annual ACM Symposium on Principles of Distributed Computing (PODC-00) », ACM Press, pages 289-298, NY, jul 16-19, 2000.

[9] P. HOMBURG, M. VAN STEEN, A. TANENBAUM. *Communication in GLOBE: an object-based worldwide operating system.* in « Proceedings of the Fifth International Workshop on Object Orientation in Operating Systems », pages 43–47, oct, 1996.

[10] A. MUTHITACHAROEN, R. MORRIS, T. M. GIL, B. CHEN. *Ivy: A Read/Write Peer-to-Peer File System.* in « Proceedings of 5th Symposium on Operating Systems Design and Implementation », 2002.

[11] D. POWELL. *Delta-4: a Generic Architecture for Dependable Distributed Computing.* Springer Verlag, 1991.

[12] R. VAN RENESSE, K. P. BIRMAN, S. MAFFEIS. *Horus: A flexible Group Communication System.* in « Communication of the ACM », number 4, volume 39, apr, 1996, pages 76-83.

Doctoral dissertations and “Habilitation” theses

[13] A. JEBALI. *Contrôle de divergence dans des environnements faiblement connectés.* Ph. D. Thesis, Université de Versailles Saint Quentin, nov, 2003.

[14] O. MARIN. *The DARX framework: Adapting Fault Tolerance for Agent Systems.* Ph. D. Thesis, Université du Havre, dec, 2003.

[15] S. PATARIN. *Pandora: support pour des services de métrologie à l’échelle d’Internet.* Ph. D. Thesis, Université Pierre et Marie Curie – Paris 6, june, 2003.

Articles in referred journals and book chapters

[16] N. DORTA, P. SENS, C. KHOURY. *Seamlessly and Coherently Locating Interesting Mirrors on the Web.* in « Journal on Systemics, Cybernetics and Informatics », number 2, volume 1, 2003.

[17] F. OGEL, G. THOMAS, A. GALLAND, B. FOLLIOT. *MVV : une Plate-forme à Composants Dynamiquement Reconfigurables — La Machine Virtuelle Virtuelle.* J.-L. GIAVITTO, editor, in « Numéro Spécial Technique et Science Informatiques (TSI) », Hermes Science, 2004, (to appear).

[18] F. OGEL, G. THOMAS, I. PIUMARTA, A. GALLAND, B. FOLLIOT, C. BAILLARGUET. *Towards Active Applications: the Virtual Virtual Machine Approach.* M. CRAUS, editor, in « New Trends in Computer Science and Engineering », series “Gh. Asachi” Technical University, Computer Engineering Department, Iasi, Romania, 2003, (to appear).

Publications in Conferences and Workshops

[19] M. BERTIER, O. MARIN, P. SENS. *Performance Analysis of Hierarchical Failure Detector.* in « Proceedings of the International Conference on Dependable Systems and Networks (DSN ’03) », IEEE Society Press, San-Francisco (USA), june, 2003.

[20] I. CHABBOUH, M. MAKPANGOU, H. B. AYED. *A Measurement System for E-Marketplaces.* in « The Sixth International Conference on Electronic Commerce Research (ICECR-6) », 2003.

[21] C. FERDEAN, M. MAKPANGOU. *A Scalable Replica Selection Strategy based on Flexible Contracts.* in « 3rd IEEE Workshop on Internet Applications (WIAPP) », San Joé (USA), 2003.

[22] O. MARIN, M. BERTIER, P. SENS. *DARX - A Framework For The Fault-Tolerant Support Of Agent Software.* in « Proceedings of the 14th IEEE International Symposium on Sofwatare Reliability Engineering (ISSRE ’03) », IEEE Society Press, Denver (USA), nov, 2003.

[23] A. MISLOVE, C. REIS, A. POST, P. WILLMANN, P. DRUSCHEL, D. WALLACH, X. BONNAIRE, P. SENS, J.-M. BUSCA, L. ARANTES. *POST: A Secure, Resilient, Cooperative Messaging System.* in « Ninth IEEE Workshop on Hot Topics in Operating Systems (HotOS-IX) », IEEE Society Press, IEEE, editor, Kaui (USA), june, 2003.

[24] F. OGEL, B. FOLLIOT, I. PIUMARTA. *On Reflexive and Dynamically Adaptable Environments for Distributed*

Computing. in « Proceedings of the 23rd International Conference on Distributed Computing Systems - Workshops (ICDCS 2003 Workshops) », Providence, Rhode Island, May, 2003.

[25] F. OGEL, G. THOMAS, B. FOLLIOT, I. PIUMARTA. *Application-Level Concurrency Management.* in « Proceedings of the Concurrent Information Processing and Computing NATO Advanced Research Workshop (CIPC 2003) », Sinaia, Romania, July, 2003.

[26] G. THOMAS, B. FOLLIOT, F. OGEL. *Jnjvm : Une PLateforme Java Adaptable pour Applications Actives.* in « Actes de la 3ème Conférence Française sur les Systèmes d'Exploitation, Chapitre français de l'ACM-SIGOPS, GDR ARP », La Colle sur Loup, France, Octobre, 2003.

Internal Reports

[27] F. L. FESSANT, S. PATARIN. *MLdonkey, a Multi-Network Peer-to-Peer File-Sharing Program.* Technical report, number 4797, INRIA, 2003, http://www.inria.fr/rrrt/rr-4797.html.

[28] S. PATARIN, M. MAKPANGOU. *On-line Measurement of Web Proxy Cache Efficiency.* Technical report, number 4782, INRIA, 2003, http://www.inria.fr/rrrt/rr-4782.html.