INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team PARIS

Programming Parallel and Distributed Systems for Large Scale Numerical Simulation Applications

Rennes - Bretagne-Atlantique

Theme : Distributed and High Performance Computing

Activity Report

2009

Table of contents

1. Team

Research Scientist

Yvon Jégou [Research Associate (CR), INRIA]
Christine Morin [Research Director (DR), INRIA, team leader from July 2009, HdR]
Thierry Priol [Research Director (DR), INRIA, team leader till June 2009, part time, HdR]

Faculty Member

Françoise André [Professor, UNIVERSITY RENNES 1, HdR]
Jean-Pierre Banâtre [Professor, UNIVERSITY RENNES 1, collaborator, HdR]
Nikos Parlavantzas [Assistant Professor, INSA, since September 2009]
Jean-Louis Pazat [Professor, INSA, HdR]
Cédric Tedeschi [Assistant Professor, UNIVERSITY RENNES 1, since September 2009]

Technical Staff

Surbhi Chitre [INRIA Associate Engineer (IA), XTREEMOS IP Project, till September 2009]
Peter Linnell [INRIA Engineer, XTREEMOS IP Project, since May 2009]
David Margery [INRIA Research Engineer (IR, ADT Aladdin)]
Pascal Morillon [UNIVERSITY RENNES 1 Engineer (IE, ADT Aladdin)]
Marko Obrovac [INRIA Engineer, XTREEMOS IP Project, since September 2009]
Cyril Rohr [INRIA Associate Engineer (IA), ADT Aladdin]

PhD Student

Boris Daix [CIFRE EDF Industrial Grant, until April 2009]
Erwan Daubert [INRIA Grant, since october 2009]
Eugen Feller [INRIA Grant, since December 2009]
Hector Fernandez [INRIA Grant, since April 2009]
Jérôme Gallard [INRIA Grant]
Guillaume Gauvrit [MENRT Grant]
Sylvain Jeuland [INRIA Grant]
André Lage [INRIA Grant]
Pierre Riteau [MENRT Grant]
Thomas Ropars [MENRT Grant]
Chen Wang [INRIA Grant, since October 2009]
Mohamed Zouari [INRIA and Brittany Regional Council Grant]

Administrative Assistant

Maryse Fouché [Team administrative assistant (TR) INRIA]
Sandrine L'Hermitte [XTREEMOS Scientific Coordinator Assistant INRIA]

2. Overall Objectives

2.1. General objectives

PARIS is a joint project of INRIA, CNRS, UNIVERSITY RENNES 1, and INSA within IRISA (UMR 6074).

In July 2009, KerData research team, headed by Luc Bougé, has been created, as a spin-off of PARIS *project-team. The research activities carried out by Luc Bougé, Gabriel Antoniu and their PhD students from January to December 2009 are all described in KerData annual activity report.*

The PARIS Project-Team aims at contributing to the programming of parallel and distributed infrastructures for large-scale numerical simulation applications. Its goal is to design operating systems and middleware to ease the use of such computing infrastructures for the targeted applications. Such applications enable the speed-up of the design of complex manufactured products, such as cars or aircrafts, thanks to numerical simulation techniques.

As computer performance rapidly increases, it is possible to foresee in the near future comprehensive simulations of these designs that encompass multi-disciplinary aspects (structural mechanics, computational fluid dynamics, electromagnetism, noise analysis, etc.). Numerical simulations of these different aspects are not carried out by a single computer due to the lack of computing and memory resources. Instead, several clusters of inexpensive PCs, and probably federations of clusters (aka. *Grids*), have to be simultaneously used to keep simulation times within reasonable bounds. Moreover, simulations have to be performed by different research teams, each of them contributing its own simulation code. These teams may all belong to a single company, or to different companies possessing appropriate skills and computing resources, thus adding geographical constraints. By their very nature, such applications will require the use of a computing infrastructure that is *both* parallel and distributed.

The PARIS Project-Team is engaged in research along five topics: *Operating System and Runtime for Clusters and Grids*, *Middleware Systems for Computational Grids*, *Large-Scale Data Management for Grids*, *Advanced Programming Models for the Grid* and *Experimental Grid Infrastructures*.

The research activities of the PARIS Project-Team encompass both basic research, seeking conceptual advances, and applied research, to validate the proposed concepts against *real* applications. The project-team is also heavily involved in managing a national grid computing infrastructure (GRID'5000) enabling large-scale experiments.

2.1.1. Parallel processing to go faster

Given the significant increase of the performance of microprocessors, computer architectures and networks, clusters of standard personal computers now provide the level of performance to make numerical simulation a handy tool. This tool should not be used by researchers only, but also by a large number of engineers, designing complex physical systems. Simulation of mechanical structures, fluid dynamics or wave propagation can nowadays be carried out in a couple of hours. This is made possible by exploiting multi-level parallelism, simultaneously at a fine grain within a microprocessor, at a medium grain within a single multi-processor PC, and/or at a coarse grain within a cluster of such PCs. This unprecedented level of performance definitely makes numerical simulation available for a larger number of users such as SMEs. It also generates new needs and demands for more accurate numerical simulation. Parallel processing alone cannot meet this demand.

2.1.2. Distributed processing to go larger

These new needs and demands, mixing high-performance and collaborative multidisciplinary works, are motivated by the constraints imposed by a worldwide economy: making things faster, better and cheaper.

2.1.2.1. Large-scale numerical simulation.

Large scale numerical simulation will without a doubt become one of the key technologies to meet such constraints. In traditional numerical simulation, only one simulation code is executed. In contrast, it is now required to *couple* several such codes together in a single simulation.

A large-scale numerical simulation application is typically composed of several codes, not only to simulate one physics, but to perform multi-physics simulation. One can imagine that the simulation times will be in the order of weeks and sometimes months depending on the number of physics involved in the simulation, and depending on the available computing resources.

Parallel processing extends the number of computing resources locally: it cannot significantly reduce simulation times, since the simulation codes will not be localized in a single geographical location. This is particularly true with the global economy, where complex products (such as cars, aircrafts, etc.) are not designed by a single company, but by several of them, through the use of subcontractors. Each of these companies brings its own expertise and tools such as numerical simulation codes, and even its private computing resources. Moreover, they are reluctant to give access to their tools as they may at the same time compete for some other projects. It is thus clear that distributed processing cannot be avoided to manage large-scale numerical applications

2.1.2.2. Resource aggregation.

More generally, the development of large scale distributed systems and applications now relies on resource sharing and aggregation. Distributed resources, whether related to computing, storage or bandwidth, are aggregated and made available to the whole system. Not only does this aggregation greatly improve the performance as the system size increases, but also many applications would simply not have been possible without such a model (peer-to-peer file sharing, ad-hoc networks, application-level multicast, publish-subscribe applications, etc.).

2.1.3. Scientific challenges of the Paris Project-Team

The design of large-scale simulation applications raises technical and scientific challenges, both in applied mathematics and computer science. The PARIS Project-Team mainly focuses its effort on Computer Science. It investigates new approaches to build software mechanisms that hide the complexity of programming computing infrastructures that are *both* parallel and distributed. Our contribution to the field can thus be summarized as follows:

combining parallel and distributed processing whilst preserving performance and transparency.

This contribution is developed along three directions.

Operating system and runtime for clusters and grids. The challenge is to design and build an operating system for clusters and grids hiding to programmers and users the fact that resources (processors, memories, disks) are distributed.

Advanced programming models for the Grid. This topic aims at contributing to study unconventional approaches for the programming of grids based on the *chemical metaphors*. The challenge is to exploit such metaphors to make the use and programming of grids more intuitive and simpler.

Experimental Grid Infrastructures. The challenge here is to be able to design and to build an *instrument* (in the sense of a large scientific instrument, like a telescope) for computer scientists involved in grid research. Such an instrument has to be highly reconfigurable and scalable to several thousand of resources.

2.2. Highlights of the year

In 2009, the project-team has the following highlight:

- The XTREEMOS European project, coordinated by the project-team (Ch. Morin) announced the second public version of the XTREEMOS Grid Operating System.

3. Scientific Foundations

3.1. Introduction

Research activity within the PARIS Project-Team encompasses several areas: operating systems, middleware and programming models. We have chosen to provide a brief presentation of some of the scientific foundations associated with them.

3.2. Data consistency

A shared virtual memory system provides a global address space for a system where each processor has only physical access to its local memory. Implementing such a concept relies on the use of complex cache coherence protocols to enforce data consistency. To allow the correct execution of a parallel program, it is required that a read access performed by one processor returns the value of the last write operation previously performed by any other processor. Within a distributed or parallel a system, the notion of the *last* memory access is sometimes only partially defined, since there is no global clock to provide a total order of the memory operation.

It has always been a challenge to design a shared virtual memory system for parallel or distributed computers with distributed physical memories, capable of providing comparable performance with other communication models such as message-passing. *Sequential Consistency* [77] is an example of a memory model for which all memory operations are consistent with a total order. Sequential Consistency requires that a parallel system having a global address space appears to be a multiprogramming uniprocessor system to any program running on it. Such a strict definition impacts on the performance of shared virtual memory systems due to the large number of messages that are required (page access, invalidation, control, etc.). Moreover Sequential Consistency is not necessarily required to correctly run parallel programs, in which memory operations to the global address space are guarded by synchronization primitives.

Several other memory models have thus been proposed to relax the requirements imposed by sequential consistency. Among them, *Release Consistency* [72] has been thoroughly studied since it is well adapted to programming parallel scientific applications. The principle behind Release Consistency is that memory accesses are (should be?) guarded by synchronization operations (locks, barriers, etc.), so that the shared memory system only needs to ensure consistency at synchronization points. Release Consistency requires the use of two new operations: *acquire* and *release*. The aim of these two operations is to specify when to propagate the modifications made to the shared memory systems. Several implementations of Release Consistency have been proposed [75]: an *eager* one, for which modifications are propagated at the time of a release operation; and a *lazy* one, for which modifications are propagated at the time of an acquire operation. These alternative implementations differ in the number of messages that needs to be sent/received, and in the complexity of their implementation [76].

Implementations of Release Consistency rely on the use of a logical clock such as a vector clock [80]. One of the drawback of such a logical clock is its lack of scalability when the number of processors increases, since the vector carries one entry per processor. In the context of computing systems that are both parallel and distributed, such as a grid infrastructure, the use of a vector clock is impossible in practice. It is thus necessary to find new approaches based on logical clocks that do not depend on the number of processors accessing the shared memory system. Moreover, these infrastructures are natively *hierarchical*, so that the consistency model should better take advantage of it.

3.3. High availability

> "A distributed system is one that stops you getting any work done when a machine you've never even heard about crashes." (Leslie Lamport)

The *availability* [73] of a system measures the ratio of service accomplishment conforming to its specifications, with respect to elapsed time. A system *fails* when it does not behave in a manner consistent with its specifications. An error is the consequence of a *fault* when the faulty part of the system is activated. It may lead to the system *failure*. In order to provide highly-available systems, *fault tolerance techniques* [78] based on redundancy can be implemented. Abstractions like *group membership*, *atomic multicast*, *consensus*, etc. have been defined for fault-tolerant distributed systems.

Error detection is the first step in any fault tolerance strategy. *Error treatment* aims at avoiding that the error leads to the system failure.

Fault treatment consists in avoiding that the fault be activated again. Two classes of techniques can be used for fault treatment: *reparation* which consists in eliminating or replacing the faulty module; and *reconfiguration* which consists in transferring the load of the faulty element to valid components.

Error treatment can be of two forms: *error masking* or *error recovery*. Error masking is based on hardware or software redundancy in order to allow the system to deliver its service despite the error. Error recovery consists in restoring a correct system state from an erroneous state. In *forward error recovery* techniques, the erroneous state is transformed into a safe state. *Backward error recovery* consists in periodically saving the system state, called a *checkpoint*, and rolling back to the last saved state if an error is detected.

A *stable storage* guarantees three properties in presence of failures: (1) *integrity*, data stored in stable storage is not altered by failures; (2) *accessibility*, data stored in stable storage remains accessible despite failures; (3) *atomicity*, updating data stored in stable storage is an all or nothing operation. In the event of a failure during the update of a group of data stored in stable storage, either all data remain in their initial state or they all take their new value.

3.4. Data management in Grids

Past research on distributed data management led to three main approaches. Currently, the most widely-used approach to data management for distributed grid computation relies on *explicit data transfers* between clients and computing servers. As an example, the *Globus* [59] platform provides data access mechanisms (like data catalogs) based on the *GridFTP* protocol. Other explicit approaches (e.g., *IBP*) provide a large-scale data storage system, consisting of a set of buffers distributed over Internet. The user can "rent" these storage areas for efficient data transfers.

In contrast, *Distributed Shared Memory* (DSM) systems provide *transparent* data sharing, via a virtual, unique address space accessible to physically distributed machines. It is the responsibility of the DSM system to localize, transfer, replicate data, and guarantee their consistency according to some semantics. Within this context, a variety of consistency models and protocols have been defined. Nevertheless, existing DSM systems have generally shown satisfactory efficiency only on small-scale configurations, up to a few tens of nodes.

Recently, *peer-to-peer* (P2P) has proven to be an efficient approach for large-scale resource (data or computing resources) sharing [81]. The peer-to-peer communication model relies on a symmetric relationship between peers which may act both as clients and servers. Such systems have proven able to manage very large and dynamic configurations (millions of peers). However, several challenges remain. More specifically, as far as data sharing is concerned, most P2P systems focus on sharing *read-only* data, that do not require data consistency management. Some approaches, like *OceanStore* and *Ivy*, deal with *mutable* data in a P2P with restricted use. Today, one major challenge in the context of large-scale, distributed data management is to define appropriate models and protocols allowing to guarantee both *consistency* of replicated data and *fault tolerance*, in *large-scale, dynamic environments*.

3.5. Component model

Software component technology [85] has been emerging for some years, even though its underlying intuition is not very recent. Building an application based on components emphasizes programming by *assembly*, that is, *manufacturing*, rather than by *development*. The goals are to focus expertise on specific domains, to improve software quality, and to decrease the time-to-market thanks to reuse of existing codes.

Component technology is being widely applied to building distributed and parallel applications. The Common Component Architecture (CCA) is a component model developed by the CCA Forum that specifically addresses the needs of the HPC community [61]. It defines a minimal set of standard interfaces that any high-performance component framework should provide to and expect from components, in order to allow them to be composed into a running application. The Grid Component Model (GCM) is a component model developed within the COREGRID Network of Excellence that targets developing grid applications [66]. GCM is defined as an extension of Fractal, a general component model for implementing, deploying, and managing complex software systems [68].

Recent component models are increasingly including support for service-oriented computing, which currently receives a lot of attention from industry and academia. Service-oriented computing promotes building flexible systems by dynamically discovering and integrating services provided over the network [83]. OSGI is a popular service-oriented Java-based component model that enables components to publish, discover and bind to services in a dynamic way [82].

3.6. Adaptability

Due to the dynamic nature of large-scale distributed systems in general, and the Grid in particular, it is very hard to design an application that fits well in any configuration. Moreover, constraints such as the number of available processors, their respective load, the available memory and network bandwidth are not static. For these reasons, it is highly desirable that an application could take into account this dynamic context in order to get as much performance as possible from the computing environment.

Dynamic adaptation of a program is the modification of its behavior according to changes of the environment. This adaptivity can be achieved in many different ways, ranging from a simple modification of some parameters, to the total replacement of the running code. In order to achieve adaptivity, a program needs to be able to get information about the environment state, to make a decision according to some optimization rules, and to modify or replace some parts of its code.

Adaptivity has been implemented by designing ad hoc applications that take into account the specificities of the target environment. For example, this was done for the Web applications access protocol on mobile networks by defining the WAP protocol [60]. A more general way is to provide mechanisms enabling dynamic self-adaptivity by changing the program's behavior. In most cases, this has been achieved by embedding the adaptation mechanism within the application code. For example, the AdOC compression algorithm [74] includes such a mechanism to dynamically change the compression level according to the available resources.

However, it is desirable to separate the adaptation engine from the application code, in order to make the code easier to maintain, and to easily change or improve the adaptation policy. This was done for wireless and mobile environments by implementing a framework [70] that provides generic mechanisms for the adaptation process, and for the definition of the adaptation rules.

Along the same idea, a new framework, based on software components, has been proposed and developed [58]. This framework is used in particular to manage parallel processes in grid environment [69], [5].

An extension of this work to cope with self management of services in cloud computing is currently under study.

3.7. Chemical programming

The chemical reaction metaphor has been discussed in various occasions in the literature. This metaphor describes computation in terms of a chemical solution in which molecules (representing data) interact freely according to reaction rules. Chemical models use the multiset as their basic data structure. Computation proceeds by rewritings of the multiset which consume elements according to reaction conditions and produce new elements according to specific transformation rules.

To the best of our knowledge, the GAMMA formalism was the first "chemical model of computation" proposed as early as in 1986 [64] and extended later [65].

A GAMMA program is a collection of reaction rules acting on a multiset of basic elements. A reaction rule is made of a condition and an action. Execution proceeds by replacing elements satisfying the reaction condition by the elements specified by the action. The result of a GAMMA program is obtained when a stable state is reached that is to say when no more reactions can take place. Here is an example illustrating the GAMMA style of programming:

$$primes = \textbf{replace}\ x, y\ \textbf{by}\ y\ \textbf{if}\ multiple(x, y)$$

The reaction $primes$ computes the prime numbers lower or equal to a given number N when applied to the multiset of all numbers between 2 and N ($multiple(x, y)$ is true if and only if x is a multiple of y). Let us emphasize the conciseness and elegance of these programs. Nothing had to be said about the order of evaluation of the reactions. If several disjoint pairs of elements satisfy the condition, the reactions can be performed in parallel.

GAMMA makes it possible to express programs without artificial sequentiality. By artificial, we mean sequentiality only imposed by the computation model and unrelated to the logic of the program. This allows the programmer to describe programs in a very abstract way. In some sense, one can say that GAMMA programs express the very idea of an algorithm without any unnecessary linguistic idiosyncrasies. The interested reader may find in [65] a long series of examples (string processing problems, graph problems, geometry problems, etc.) illustrating the GAMMA style of programming and in [63] a review of contributions related to the chemical reaction model. Later, the idea was developed further into the CHAM [67], the P-systems [84], etc. Although built on the same basic paradigm, these proposals have different properties and different expressive powers.

The γ-calculus [62] is an attempt to identify the basic principles behind chemical models. It exhibit a minimal chemical calculus, from which all other “chemical models” can be obtained by addition of well-chosen features. Essentially, this minimal calculus incorporates the γ-reduction which expresses the very essence of the chemical reaction, and the associativity and commutativity rules which express the basic properties of chemical solutions.

4. Application Domains

4.1. Application Domains

The project-team research activities address scientific computing and specifically numerical applications that require the execution of several codes simultaneously (code-coupling). This kind of applications requires both the use of parallel and distributed systems. Parallel processing is required to address performance issues. Distributed processing is needed to fulfill the constraints imposed by the localization and the availability of resources, or for confidentiality reasons. Such applications are being experimented within contracts with the industry or through our participation to application-oriented research grants.

5. Software

5.1. Kerrighed

Participants: Marko Obrovac, Christine Morin, Pierre Riteau, Eugen Feller.

Contact: Christine Morin, `Christine.Morin@irisa.fr`

URL: http://www.kerrighed.org/

Status: Registered at APP, under Reference `IDDN.FR.001.480003.006.S.A.2000.000.10600`.

License: GNU General Public License (GPL) version 2. KERRIGHED is a registered trademark.

Presentation: KERRIGHED is a *Single System Image* (SSI) operating system for high-performance computing on clusters. It provides the user with the illusion that a cluster is a virtual SMP machine. KERRIGHED is based on Linux which is slightly patched and extended with a kernel module. It is Posix compliant. Legacy sequential or parallel applications running on Linux can be executed without modification on top of KERRIGHED. KERRIGHED (version V2.4.1) includes around 40,000 lines of code (mostly in C). It involved more than 250 persons-months. Professional support is provided by Kerlabs http://www.kerlabs.com, a spin-off from PARIS project-team created in 2006. KERRIGHED is used in the cluster flavour of XTREEMOS Grid operating system. In 2009, we have performed an extensive performance evaluation of the following KERRIGHED sub-systems and mechanisms: kDFS distributed file system [29], [79], global scheduler and checkpointing mechanisms. We have improved the checkpoint/recovery mechanisms to support call-backs, incremental checkpointing and to integrate with the XTREEMOS XTREEMGCP service in charge of the reliable execution of Grid applications [52]. We have further automatized the installation and configuration process of KERRIGHED when used in the framework of XTREEMOS system.

5.2. Ange

Participants: Christine Morin, Boris Daix.

Contact: Christine Morin, `Christine.Morin@inria.fr`

URL: http://www.irisa.fr/paris/software/samuraaie/

Status: Prototype.

License: GPL-2.

Presentation: ANGE (Adage New GEneration) is a proof-of-concept of the ODD/SAMURAAIE automatic deployment architecture. Developped with Python, ANGE is a deployment engine that can be called either by application middleware or directly by its user. It already features abstractors that provide full persistence for manipulated abstractions, managers that provide a complete decision making mechanism for the deployment process, and translators and connectors for YAML, CORBA, POSIX, and SSH technologies. This implementation has allowed to show how SALOME, a component-based, supervised numerical simulation platform, can use ODD/SAMURAAIE to automatically deploy its simulations on high-performance computing infrastructures.

5.3. Dynaco

Participants: Françoise André, Jean-Louis Pazat.

Contact: Jean-Louis Pazat, `Jean-Louis.Pazat@irisa.fr`

URL: http://dynaco.gforge.inria.fr/

Status: Version 0.2 is available.

License: GNU Lesser General Public License (LGPL) version 2.1.

Presentation: DYNACO (*Dynamic Adaptation for Components*) is a framework that helps in designing and implementing dynamically adaptable components. This framework is developed by the PARIS Project-Team. The implementation of DYNACO is based on the *Fractal Component Model* and its formalism.

5.4. Vigne

Participants: Christine Morin, Thomas Ropars.

Contact: Christine Morin, `Christine.Morin@inria.fr`

URL: http://www.irisa.fr/paris/software/vigne/index_html#software

Status: Prototype

License: GNU General Public License (GPL).

Presentation: VIGNE is a prototype of a grid-aware operating system for grids, whose goal is to ease the use of computing resources in a grid for executing distributed applications. VIGNE is made up of a set of operating system services based on a peer-to-peer infrastructure. This infrastructure currently implements a structured overlay network inspired from *Pastry* and an unstructured overlay network inspired from *Scamp* for join operations. On top of the structured overlay network, a transparent data-sharing service based on the sequential consistency model has been implemented. It is able to handle an arbitrary number of simultaneous reconfigurations. An application execution management service has also been implemented including resource discovery, resource allocation, and application monitoring services. In 2009, we have developed the SEMIAS framework that provides stateful services with high availability and self healing based on active replication on top of a structured overlay. The VIGNE application execution management service has been made highly available and self-healing using the SEMIAS framework. The VIGNE prototype has been developed in C and includes around 30,000 lines of code. This prototype has been coupled with a discrete-event simulator.

5.5. XtreemOS

Participants: Surbhi Chitre, Jérôme Gallard, Yvon Jégou, Sylvain Jeuland, Christine Morin, Pierre Riteau, Marko Obrovac, Peter Linnell.

Contact: Christine Morin, `Christine.Morin@inria.fr`

URL: http://www.xtreemos.eu, http://gforge.inria.fr/projects/xtreemos

Status: Version 2.0

License: GPL-2/BSD depending on software packages composing the system

Presentation: XTREEMOS is a prototype of a Grid Operating system based on Linux with native support for virtual organizations. Three flavours of XTREEMOS are developed for individual PCs, clusters and mobile devices. XTREEMOS has been developed by the XTREEMOS consortium. The second public version of XTREEMOS (PC and cluster flavours) has been released in November 2009. XTREEMOS has been demonstrated at EuroPar 2009, Delft, The Netherlands, in August 2009 (P. Linnell), at the XTREEMOS summer school, Oxford, UK in September 2009 (P. Linnell), SC'09, Portland, USA (Y. Jégou, P. Linnell, Ch. Morin) in November 2009. XTREEMOS software is a set of services developed in Java, C++ and C. XTREEMOS cluster version leverages KERRIGHED single system image operating system. A permanent testbed composed of computers provided by several XTREEMOS partners has been setup for experimentation and demonstration purposes.

5.6. Saline

Participants: Jérôme Gallard, Christine Morin.

Contact: Christine Morin, `Christine.Morin@inria.fr`

Status: Version V1.0 (experimental)

License: GPL-2

Presentation: `Saline` (former called `VMdeploy`) is a generic framework to deploy and manage encapsulated user jobs in virtual machines (VMs) at grid level by moving them from one site to another transparently for the encapsulated jobs [36]. Moreover, `Saline` is non-intrusive and can be used with any non-modified Grid resource management systems (RMSs).

`Saline` deploys and configures a set of VMs according to the user needs. Then, periodically, `Saline` takes snapshots of the running VMs and saves them on a dedicated node in an efficient way [32]. In addition, `Saline` checks the status of the running VMs. If something wrong happens *i.e.* one or more VMs have failed due for instance to a node failure or to the arrival of a higher priority job, `Saline` redeploys the set of VMs from the latest snapshot taken on new available resources provided by the Grid RMS. Thanks to `Saline`, the redeployment of the snapshot is done in a transparent way from the encapsulated job point of view.

In its current implementation, `Saline` is programmed in `bash` and `C`. It uses `Libvirt` in order to create, to snapshot, and to restart the VMs. It means that `Saline` can deploy and manage KVM and XEN VMs or any other VMs usable with `Libvirt`. In addition, the architecture of `Saline` is very modular in order to have a clear and easily extensible code.

5.7. HOCL

Participants: Thierry Priol, Cédric Tedeschi, Chen Wang.

Contact: Thierry Priol, `Thierry.Priol@inria.fr`

Status: V2.3

License: GPL-2

Presentation: HOCL (Higher Order Chemical Language) is a chemical programming language based on the chemical metaphor presented before (see Section 3.7). It was developed for several years within the PARIS team. Within HOCL, following the chemical metaphor, computations can be regarded as chemical reactions, and data can be seen as molecules which participate in these reactions. If a certain condition is held, the reaction will be triggered, thus continuing until it gets inert: no more data can satisfy any computing conditions. To realize this program paradigm, a multiset is implemented to act as a chemical tank, containing necessary data and rules. An HOCL program is then composed of two parts: *chemical rule definitions* (reaction rules) and *multiset definition* (data). More specifically, HOCL provides the high order: reaction rules are molecules that can be manipulated like any other molecules. In other words, HOCL programs can manipulate other HOCL programs.

An HOCL compiler was developed using java to execute some chemical programs expressed with HOCL. This compiler is based on the translation of HOCL programs to java code. As a support for service coordination and service adaptation (refer to Section 6.4), we recently extended the HOCL compiler with the following features:

- Possibility of communication between two multisets
- Support for the communication between two different chemical programs
- Support for passing chemical rules between two multisets
- Support for the dynamic creation and addition of chemical rules at runtime

These aspects will allow to have a software support to experiment the chemical implementation of service coordination.

6. New Results

6.1. Introduction

Research results are presented according to the scientific challenges of the PARIS Project-Team. The activities of Luc Bougé, Gabriel Antoniu and their PhD students are presented in KerData annual activity report.

6.2. Operating system and runtime for clusters and grids

6.2.1. Cluster operating systems

Participants: Marko Obrovac, Christine Morin, Eugen Feller.

6.2.1.1. Evaluation of LinuxSSI single system image operating system for clusters

In 2009, we have carried out an extensive performance evaluation of LinuxSSI (XtreemOS cluster flavour foundation layer based on KERRIGHED Linux based single system image operating system). In particular, we have evaluated the kDFS distributed parallel file system, the global scheduling policy and the checkpointing mechanisms [50].

6.2.1.2. Energy management in clusters

In 2009, we have initiated a study on energy consumption management in clusters. This work is carried out in the framework of the Eco-grappe ANR project (PhD thesis of Eugen Feller). The objective of this work is to be able to adapt the cluster configuration (hardware parameters, number of nodes) to the actual workload in order to save energy. Experimentations will be carried out with the KERRIGHED open source cluster operating system.

6.2.2. Grid operating systems

Participants: Surbhi Chitre, Marko Obrovac, Jérôme Gallard, Yvon Jégou, Sylvain Jeuland, Peter Linnell, Christine Morin, Pierre Riteau, Thomas Ropars.

6.2.2.1. Access Control and Interactive Jobs

XtreemOS aims to provide to grid users an interface similar to their usual Linux desktop interface: application are run on the grid as if they were executed on the local desktop. In order to provide this interface, XTREEMOS must provide means to support single-sign-on (no need to authenticate each time an application is run), delegation (applications running on the grid have the same capabilities as applications running on the desktop) and interactive applications (the user can interact with his application through ttys and his desktop display, grid applications can participate to pipes, ...). The single-sign-on and delegation system proposed by Yvon Jégou for XTREEMOS [40] has been selected by the security work package members and is currently being implemented. Yvon Jégou is also in charge of providing support for interactive applications. The proposed mechanism [41] has been implemented and is available since the 2.0 release of XTREEMOS.

6.2.2.2. Dynamic Virtual Organizations and Execution of Coordinated Services

A key feature of XTREEMOS is its support for Virtual Organizations (VOs). The current XTREEMOS prototype does not properly address dynamic VOs. VOs are dynamic in a number of directions: addition and removal of users and resources, creation and deletion of attributes, addition and removal of user attributes, generation and invalidation of identity and attribute certificates, automatic VO generation when a new project is set up, and VO federation. Moreover, the execution of service-based applications in the XTREEMOS framework has not been addressed when designing the first version of the system.

In 2009, we have designed and implemented the XChor functionality which permits users to communicate, exchange data and run jobs in a collaborative workflow. The multi-party aspect of choreography (in opposition to orchestration) is privileged because no centralized leader controls the execution. In our case a choreography becomes a peer-to-peer collaboration of users and accomplishes a common goal where the user behaviors are defined by job dependencies and ordered message exchanges. An XTREEMOS choreography -XChor- is executed in a Dynamic Virtual Organization [39].

Since the long-lived VOs do not permit the building of XChors, we have introduced the Dynamic Virtual Organization (DVO) concept to support short-lived multi-user collaborations. A Dynamic Virtual Organization is generated on-the-fly when partners from different administration domains decide to run a choreography. Once all users have been registered and have retrieved their DVO credentials, the XTREEMOS choreography is triggered and the collaborative jobs are submitted to XTREEMOS Grid system resources according to the defined workflow.

6.2.2.3. A Checkpointing Service for the Grid

XTREEMGCP [28] is a service of the XTREEMOS grid system that provides grid applications with fault tolerance. It is able to apply different fault tolerance strategies and to make use of the various kernel checkpointers available on the grid nodes. In 2009, we have implemented a new kernel checkpointer exploiting OpenVZ features to suspend, checkpoint and restart jobs executed on individual PC nodes and integrated it in the XTREEMGCP service. We have also improved the LinuxSSI checkpointer, providing call backs, in order to integrate it in the XTREEMGCP service. All these new funtionalities have been integrated in the second XTREEMOS release. Eugen Feller in the context of his Master internship has also implemented and integrated an independent checkpointing protocol in the XTREEMGCP service allowing to checkpoint/restart distributed applications executed on heterogeneous Grid nodes (PC, clusters). This work demonstrates the genericity of the XTREEMGCP service that is able to exploit different kernel checkpointers (BLCR, LinuxSSI checkpointer, OpenVZ based checkpointer) and to drive different checkpointing protocols for distributed applications (coordinated checkpointing, independent checkpointing). This work has been carried out in close collaboration with the Heinrich-Heine Universitaet Duesseldorf, Germany (Michael Schoettner's group). Future work includes the integration of a Grid application checkpointing protocol based on the O2P protocol described below and to exploit checkpoint/restart functionalities to migrate jobs in a Grid.

6.2.2.4. Fault Tolerance for Message Passing based Applications

O2P is an optimistic message logging protocol that targets large scale message passing applications and that has been implemented in the Open MPI Library. O2P is based on active optimistic message logging, a new message logging strategy that makes it more scalable than existing optimistic message logging protocols [30].

The scalability limit of O2P is the event logger, i.e. the centralized process used to manage logging on stable storage. This centralized event has been proposed in previous works on message logging protocols. In 2009, we have proposed a new event logger which is completely distributed to be scalable. This event logger makes use of the memory of the computation nodes to implement stable storage. Experiments show that this new event logger makes O2P more scalable than a centralized event logger [14].

A new message logging model has been defined in the context of the MPI standard, that reduces the number of events that have to be logged on stable storage by a message logging protocol. In that context, we have shown in collaboration with Aurélien Bouteiller from the University of Tennessee, that the main cost of message logging protocols is not due to event logging on stable storage anymore, but due to sender-based message logging [24].

6.2.2.5. High Availability for Grid Services

To provide high availability and self-healing for stateful services, we have proposed a new framework called Semias, which is based on the combination of peer-to-peer techniques with active replication. Active replication makes services highly available [14]. We have designed and implemented a software stack comprising of a failure detector, a consensus, an atomic broadcast, a group membership protocols and a reconfiguration service. In the framework of Sébastien Gillot's internship, we have validated the implementation of the consensus protocol (Paxos) and of the atomic broadcast protocol, using the Splay simulator. A tool to automatically deploy the Spray framework on Grid'5000 has been developed. One of the main challenge in service replication is to handle reconfigurations. In the context of Stefania Costache's Master internship, we have proposed a group monitoring layer which is in charge of gathering monitoring information about grid nodes to take appropriate reconfiguration decisions based on some safety conditions. To our knowledge, Semias is the first implementation of atomic broadcast on top of a structured peer-to-peer overlay. The peer-to-peer overlay provides fault tolerant routing mechanisms and makes replication completely transparent for the clients. Thus existing services can be replicated using Semias with very little modifications limited to the service state transfer. To validate our approach, we have used Semias to make services of the Vigne grid system highly available and self-healing. Experiments run on the Grid'5000 testbed show that Semias can replicate services with a very small overhead and can efficiently and automatically handle failures [34].

6.2.2.6. Scalable failure information base

A very important part in ensuring reliability in distributed systems is failure detection, being able to accurately determine when a node has failed and to select the best nodes for a job. In the framework of Catalin Leonardu's internship we have proposed a storage system for nodes' failure history information. This solution aims at increasing the reliability of a distributed system by providing failure information which is as accurate as possible, even in a very dynamic system. Evaluations made through simulation with Splay show the scalability of the system.

6.2.3. Federated Virtualized Infrastructures

Participants: Jérôme Gallard, Yvon Jégou, Christine Morin, Thierry Priol, Pierre Riteau.

6.2.3.1. Virtual infrastructure management

In 2009, we have further investigated the design of systems allowing to share resources in a peer-to-peer manner [37], [16]. System virtualization is an enabling technology as it allows to decouple the software from the hardware and avoids the deployment of application specific systems and libraries on the infrastructure nodes.

Based on our preliminary work done on the VMdeploy environment, we have built the `Saline` system that allows the management of applications run in virtual machines Grid-wide. `Saline` is now able to manage both regular and best efforts jobs [36] [71]. It has been interfaced with the OAR resource reservation system and experimented on Grid'5000 platform. In particular, we have designed and implemented a service managing virtual machines IP addresses in a Grid.

We have also studied in 2009 the dynamic extension of the XTREEMOS operating system in order to take advantage of cloud computing infrastructures. In the framework of Eliana-Dina Tirsa's internship we have designed and implemented a service enabling the automatic deployment of XTREEMOS system in a set of virtual machines provisioned from a Nimbus Cloud. Thus, we are now able to extend a Grid infrastructure with resources from an infrastructure as a service cloud. In future work, we will make the proposed service more generic and interface it with various kinds of IaaS clouds. We also plan to study resource management policies in the context of federated virtualized infrastructures.

In the framework of the SER-OS associated team, we have worked on the design and the implementation of a novel management tool for managing virtualized infrastructures comprising of high performance computing clusters, massively parallel processing (MPP) systems, and grids [38]. This novel management tool integrates three main concepts: (i) Virtual System Environments (VSEs) describing the application requirements in terms of software configuration (ii) Virtual Organizations (VOs) defining sets of resources shared among users communities, and (iii) Virtual Platforms (VPs) describing the application requirements in terms of hardware platform.

6.2.3.2. Virtual machine migration

In relation with the previously described research activities, another research direction focuses on the management of sets of virtual machines aiming at improving migration and storage of virtual machines in grid/cloud computing environments.

Best efforts jobs running in a set of virtual machines need to be suspended and stored to permanent storage to free a set of nodes in order to execute higher priority jobs. We have developed the Kget+ tool enabling fast removal and storage of a set of virtual machines from a set of nodes without contention on the target storage system [32]. This tool is exploited in the `Saline` system.

We have investigated the use of distributed content addressing to enable efficient live migration of virtual machines over wide-area networks. We designed a customized live migration protocol that only sends cryptographic hashes of the virtual machine to the destination node, which results in much lesser network traffic. The target node then leverages a distributed hash table on the remote site to find VM pages that are already present in the local network. VM pages that cannot be found are requested from the original node.

A prototype has been implemented as a modification of the QEMU/KVM hypervisor. This prototype is currently being evaluated on the Grid'5000 platform.

In the context of Djawida Dib's Master internship, we have initialized work on network transparent live migration, allowing virtual machines to migrate without any impact on their network connections. This will enable parallel jobs relying on network communications (e.g. MPI programs) to live migrate to a remote site.

6.2.3.3. XtreemOS Release and Deployment Tools

Yvon Jégou and Peter Linnell, as release manager, coordinated the production and the testing of the second integrated version of XTREEMOS Grid operating system (XtreemOS V2.0 version), publicly released in November 2009 (http://www.xtreemos.eu/software). We have contributed to the XTREEMOS admin and user guides [57], [56]. The permanent geographically distributed testbed made up of several computers provided by different XTREEMOS partners has been updated with the new XTREEMOS release and used for testing and demonstrating the XTREEMOS prototype.

In 2009, we have developed a set of tools and environments to facilitate the deployment of the XTREEMOS system on various infrastructures. We have designed and implemented a configuration tool to facilitate the deployment of XTREEMOS system on a Grid made up of physical or virtual machines. Moreover, pre-configured sets of virtual machines for KVM and VirtualBox hypervisors have been produced, documented and made available for the XTREEMOS consortium and the open source community (http://www.xtreemos.eu/software). Tools to automatically deploy XTREEMOS PC and cluster flavours on Grid'5000 platform have been developed and made available to XTREEMOS consortium.

6.3. Middleware systems for computational grids

6.3.1. Application deployment on computational Grids

Participants: Boris Daix, Christine Morin.

Computing is an important lever in scientific, technical, and industrial development. However using high performance systems remains a complex activity. Indeed, allocating some computing resources for a given application, installing it on the allocated resources, and executing it is a difficult process for most users of these systems. Automating this process, called deployment, allows to better separate the preoccupations of the system users, of the application developers, and of the infrastructure administrators. This automation also allows to aim at better performance for the applications. Based on a multi-level abstraction model, ODD/SAMURAAIE (On Demand Deployment/System Abstraction Model for User, Resources, And Applications (Actions on, Instance of, and Events from)), our contribution is an automatic deployment architecture [13]. In 2009, we have finalized the ANGE proof of concept implementation this architecture. Three significant deployment cases have been automatized thanks to the ODD/SAMURAAIE automatic deployment architecture. This architecture is based on a multi-level system abstraction model (SAMURAAIE) and provides a set of concurrent actors to manage on-demand deployment (ODD) for high-performance computing applications. The deployment cases that have been studied are a parallel application on a remote computer, a master-slave application on a remote computer cluster, and a workflow application on a remote computer grid. In the first case, the application is featured as a directory of compiled files and the computer is multi-processor. In the second case, the application is featured as a compiled file archive and the number of available computers in the cluster changes dynamically. In the third case, the workflow application is featured as a source code archive, contains three sequential codes and a parallel code, and the grid contains two heterogeneous computer clusters. With their growing complexity, the automation of these cases has shown how pertinent are the system abstraction and the on-demand deployment models of the ODD/SAMURAAIE architecture.

A SALOME numerical simulation has been deployed thanks to ANGE and the ODD/SAMURAAIE automatic deployment architecture. SALOME is a component-based, supervised numerical simulation platform. It currently requires that its user allocates some computers, installs SALOME components on them, and declares these "SALOME resources" to the platform itself, all that by hand. After having analyzed this deployment approach, ODD translators and connectors have been developed so that ANGE can automatically deploy SALOME simulations on remote infrastructures. A preliminary experiment has shown how SALOME can delegate, to ANGE with SALOME support, the deployment of a simple numerical simulation. Future work will improve the internal deployment approach of SALOME so that it can benefit of all automatic deployment features that ANGE offers.

This work is carried out in collaboration with Christian Perez from GRAAL project-team. Since April 2009, this work is pursued at EDF R&D where Boris Daix has been recruited.

6.3.2. Adaptation for data management

Participants: Françoise André, Mohamed Zouari.

The usage of context-aware data management in mobile environments has been investigated by Françoise André in collaboration with Mayté Segarra and Jean-Marie Gilliot from Telecom Bretagne Brest (previously known as ENST Bretagne). A context-aware data replication and consistency system that adapts dynamically to changes in the environment has been proposed, based on the use of the DYNACO framework. This work has been supported by a contract (*ReCoDEM*) between ENST Bretagne and Orange Labs (previously known as France-TélécomR&D).

In the *ReCoDEM* project, the distributed aspects of the adaptation system have not been thoroughly investigated. Therefore, a new subject is launched since October 2007 (with M. Zouari as PhD student) to propose a generic distributed adaptation framework. This work focuses on data management in grid and mobile environments; an ambient assisted living application illustrates the approach. Two architectural models have been proposed, one for distributed data management and the other for distributed dynamic adaptation. The models have been implemented as two component-based frameworks which are currently under evaluation. This work is realized in the context of the *ALORAD* project (Architecture LOgicielle pour la Réplication Adaptative de Données), financed by the Brittany council. Mayté Segarra from Telecom Bretagne Brest is co-adviser for the PhD thesis of M. Zouari.

6.3.3. Adaptation for Service-Oriented Architectures

Participants: Françoise André, Guillaume Gauvrit, Erwan Daubert, André Lage, Nikos Parlavantzas, Jean-Louis Pazat, Chen Wang.

Service-Oriented Computing is a paradigm that is rapidly spreading in all application domains and all environments - grids, clusters of computers, mobile and pervasive platforms. The necessity of dynamic self-adaptation of services to satisfy the different users needs on constantly changing environments is evident.

The following works take place in the context of the S-CUBE European Network of Excellence

6.3.3.1. Services adaptation in distributed and heterogeneous systems

Based on our experience on components adaptation, Françoise André, Guillame Gauvrit and Erwan Daubert are now studying the problems of services adaptation in distributed and heterogeneous systems. This work takes place in the context of the S-CUBE European Network of Excellence and covers different aspects such as structural, behavioral and environmental adaptation, distributed decision and planification of adaptation actions, adaptive allocation of resources for services.

6.3.3.2. Local Service Adaptation

Jean-Louis Pazat and André Lage are studying how to apply dynamic adaptation principles at the level of one single service running on a Grid infrastructure. We are building a prototype based on the OSGi component framework, the XTREEMOS (XO SAGA) interface and the Wildcat monitoring tool.

We intend to define explicit links between QoS of OSGi services and Grid resource utilization in order to be able to optimize resource usage while conforming to some SLA.

6.3.3.3. Dynamic Adaptation of Chemical services

Jean-Louis Pazat and Chen Wang are studying how to express and implement dynamic adaptation of services within the context of chemical programming. We are also working on extensions of the HoCl compiler in order to remove some limitations and add new dynamic features.

6.4. Advanced programming models for the Grid

Participants: Jean-Pierre Banâtre, Hector Fernandez, Jean-Louis Pazat, Thierry Priol, Cédric Tedeschi, Chen Wang.

In the context of web services, we have started to study how to take advantage of the good properties of the chemical computing paradigm (implicit parallelism, autonomic behavior) to express service coordination. For this purpose, we are defining the S-HOCL language, since the basic HOCL language does not support service coordination, and we are extending the HOCL compiler in several ways (refer to Section 5.7).

In collaboration with STFC (UK) under the framework of the CoreGRID Network of Excellence, we are studying the security engineering of distributed systems, such as Grids, when following the chemical-programming paradigm. We have analysed how to model secure systems using HOCL. Since our emphasis is on modularity, we advocate the use of aspect-oriented techniques, where security is seen as a cross-cutting concern impacting the whole system. We showed how HOCL can be used to model Virtual Organisations (VOs), exemplified by a VO system for the generation of digital products. We also develop security patterns for HOCL, including patterns for security properties such as authorisation, integrity and secure logs [3].

As mentioned earlier (refer to Section 6.3.3.3), we are also studying how to express and implement service adaptation through a chemical approach.

Finally, we started to study the development of the prototype of a middleware system for the distributed execution of chemical programs (targeted for large scale platforms).

7. Contracts and Grants with Industry

7.1. EDF Contract 2 (2006-2009)

Participants: Boris Daix, Christine Morin.

The collaboration with EDF R&D aims at improving the dynamic deployment of scientific code-coupling applications on cluster federations, taking into account their execution constraints. In 2009, we have integrated ANGE automatic deployment engine implementing the ODD/SAMURAAIE architecture and model with the SALOME numerical simulation platform [13].

8. Other Grants and Activities

8.1. Regional grants

8.1.1. Brittany Regional Council

The Brittany Regional Council provides half of the financial support for the PhD thesis of Mohamed Zouari. This support amounts to a total 14,000 Euros/year.

The subject of M. Zouari's PhD thesis relates to the *ALORAD* project regarding the adaptation of data management in distributed applications.

8.1.2. Rennes Metropole

Jérôme Gallard has been awarded a grant from Rennes Metropole for his three-month internship at ORNL (October 2009 - January 2010).

8.2. National grants

8.2.1. ANR WP: ANR White Program

8.2.1.1. ANR WP AutoChem Project

Participants: Jean-Pierre Banâtre, Thierry Priol.

This project aims at investigating and exploring an unconventional approach, based on chemical computing, to program complex computing infrastructures, such as Grids and real-time deeply-embedded systems. It is a 3-year project which started in December 2007.

8.2.2. INRIA ADT

8.2.2.1. ALADDIN-G5K

Participants: Yvon Jégou, David Margery, Pascal Morillon, Thierry Priol, Cyril Rohr.

The PARIS project-team coordinated in 2008 a follow-up proposal to the Grid'5000 project of the ACI GRID. This proposal aims at the construction of a scientific instrument for experiments on large-scale parallel and distributed systems, building on the Grid'5000 platform. This proposal was accepted and officially started June 1st. It structures INRIA's leadership role as the institute is present in 8 if the 9 Grid'5000 sites distributed across France.

Frédéric Desprez is the director of this ADT, Franck Cappello the scientific director, and David Margery the technical director. An executive committee, where each of the 10 project-teams supporting Grid'5000 in the 8 research centers is represented, meets every month. It gives recommendations to the directors on scientific animation, access policy to the instrument as well as for the hardware and software development according to the resources devoted to this ADT. Yvon Jégou represents INRIA RENNES – BRETAGNE ATLANTIQUE in this executive committee.

The technical team is now composed of 12 engineers, of which 3 are from the PARIS project-team (David Margery, Pascal Morillon, Cyril Rohr). This technical team is structured in a sysadmin team, managing the instrument, and a development team building the tools to build, execute and analyze experiments.

30 project-teams not including the 10 mentioned above are using or have shown interest in using such an instrument.

8.2.3. ANR TLOG: ANR Program on Software Technologies

8.2.3.1. ANR TLOG NeuroLog Project

Participant: Yvon Jégou.

The *NeuroLog* consortium (*Software technologies for integration of process, data and knowledge in medical imaging*) is targeting software technologies in medical domains for large scale management of data, knowledge and computation: management and access of partly structured data, heterogeneous and distributed in an open environment; access control and protection of private medical data; control of workflows implied in complex computing process on grid infrastructures; extraction and quantification of relevant parameters for different pathologies.

8.2.3.2. ANR ARPEGE Eco-Grappe Project

Participants: Christine Morin, Eugen Feller.

We are involved in the EcoGrappe consortium with EDF R&D and Kerlabs. The 3 year Eco-Grappe project started in December 2008 and funded under the ANR ARPEGE programme aims at the management of energy consumption is clusters. Our work, started in December 2009, relates to the design and implementation of energy saving resource management policies in clusters running a single system image operating system such as KERRIGHED. Eugen Feller's PhD grant is funded under this project.

8.2.3.3. ANR COSINUS COOP Project

Participants: Christine Morin, Yvon Jégou.

We are involved in the COOP consortium with INRIA Grenoble (coordinator), INRIA Bordeaux, IRIT and EDF R&D. COOP is funded under the ANR COSINUS programme and is a 3 year project started in December 2009. This project aims at studying the interactions between programming environments or runtimes and resource management systems for efficient execution of HPC applications. We will perform experiments with Vigne and XTREEMOS Grid systems and the SALOME platform for numerical simulation. A post-doctoral (18 months) and a research engineer (18 months) positions are funded.

8.3. European grants

8.3.1. XtreemOS IP Project

Participants: Surbhi Chitre, Marko Obrovac, Jérôme Gallard, Yvon Jégou, Sylvain Jeuland, Peter Linnell, Sandrine L'Hermitte, Christine Morin, Thierry Priol, Thomas Ropars.

Christine Morin is the *Scientific Coordinator* (SCO) of the XTREEMOS Integrated Project (IP) that addresses Strategic Objective 2.5.4 *Advanced Grid Technologies, Systems and Services*, Focus 3 on *Network-centric Grid Operating Systems* as described in the IST 2006 Work Programme. XTREEMOS involves 19 academic and industrial partners from 7 European countries plus China. The XTREEMOS project aims at the design, implementation, evaluation and distribution of an open source Grid operating system with native support for virtual organizations and capable of running on a wide range of underlying platforms, from clusters to mobiles. The approach we propose in this project is to investigate the construction of a new Grid OS, XTREEMOS, based on the existing general-purpose OS Linux [6]. *Yvon Jégou* leads the WP4.3 Work-Package, aiming at setting up XTREEMOS testbeds. The GRID'5000 experimental grid platform is used as a large-scale testbed by XTREEMOS partners. Christine Morin leads WP1.1, Project management, WP2.1, Virtual Organization support in Linux, WP2.2 Federation management and WP5.3, Collaboration with other IST Grid-related projects. Thierry Priol is a member of the *Scientific Advisory Committee*.

8.3.2. S-Cube NoE Project

Participants: Françoise André, Erwan Daubert, Guillaume Gauvrit, André Lage, Nikos Parlavantzas, Jean-Louis Pazat, Chen Wang.

S-Cube is a European Network of Excellence in Software Services and Systems. Its goal is to establish an integrated, multidisciplinary, vibrant research community. This will enable Europe to lead the software-services revolution, thereby helping shape the software-service based Internet which is the backbone of our future interactive society.

We are now leader of the infrastructure workpackage which aims at defining specifications of the infrastructure for building adaptable and self-* service based applications.

In this area we are working on adaptation at the level of one single service (André Lage's PhD thesis) and on distributed and multilevel adaptation. The different phases of the adaptation process are studied: in his PhD thesis Guillaume Gauvrit focuses on the problem of the decision phase in a distributed and multi-levels context, whereas Erwan Daubert's thesis tackles the problem of the planification and implementation of adaptation actions for distributed heterogeneous Service-Oriented platforms. We are also investigating the use of non-conventional programming paradigms for service composition and execution (PhD thesis of Chen Wang).

8.3.3. IC0804 - Energy efficiency in large scale distributed systems

Participants: Françoise André, Jean-Louis Pazat.

This COST Action will propose realistic energy-efficient alternate solutions to share IT distributed resources. As large scale distributed systems gather and share more and more computing nodes and Storage resources, their energy consumption is exponentially increasing. While much effort is nowadays put into hardware specific solutions to lower energy consumptions, the need for a complementary approach is necessary at the distributed system level, i.e. middleware, network and applications. The Action will characterize the energy consumption and energy efficiencies of distributed applications.

8.4. International Collaborations

8.4.1. North-America

8.4.1.1. SER-OS associated team

We have launched in 2009 the SER-OS associated team (http://www.irisa.fr/paris/ser-os/index.html) with the SRT (*System Research Team*) team, ORNL (headed by Stephen Scott) on virtualization and resilience in large scale distributed infrastructures (clusters, supercomputers, grids). A workshop has been organized at INRIA Rennes - Bretagne Atlantique on May 13-15, 2009. The scientific leader of the SER-OS associated team is Christine Morin.

8.4.1.2. Informal collaboration with University of Chicago/Argonne

We have initiated a collaboration with Kate Keahey, head of the research group developing Nimbus IaaS Cloud technology. We have jointly worked on dynamic resource provisioning from Clouds in Grids. Experiments have been conducted using Nimbus and XTREEMOS technologies. We are involved in a PIRE proposal on federated virtualized infrastructures submitted by Kate Keahey.

8.4.2. Middle-East, Asia, Oceania

8.4.2.1. UNCONV associated team

The UNCONV associated team gathers together the PARIS team from INRIA Rennes Bretagne Atlantique, the POP-ART team from INRIA Grenoble - Rhône-Alpes and the Insititute of Computing Technology from the Chinese Academy of Sciences. Thierry Priol is the scientific leader of the UNCONV team.

UNCONV builds upon the observation that reliability of Grid Infrastructures do not yet reach the level at which it can be used for critical applications and programming Grids is very similar with what computer programmers faced before programming languages came in the light. The objective of UNCONV is to investigate disruptive and unconventional approaches in the design of Grid systems and Grid programming paradigms. UNCONV proposes to study how existing open-source operating systems could be extended to facilitate the design of reliable Grid infrastructures and to investigate a new programming paradigm based on the chemical metaphor.

8.4.3. Africa

8.4.3.1. Tunisia

This project is a DGRSRT/INRIA project between the "Ecole Nationale d'Ingénieurs de SFAX" (ENIS), Tunisia, and INRIA RENNES – BRETAGNE ATLANTIQUE (EPI VERTECS and PARIS). The project aims at defining a model-based approach for the test of reactive systems. It suggests studying software validation in the context of sensors networks, control-command systems and adaptive applications. A generic methodology is proposed to ensure distributed systems validation. A set of tools will be developed to support the methodology and applied to different applications. Exchanges of researchers, PhD and Master students foster the research collaboration, ideas exchange and the adoption of the results.

9. Dissemination

9.1. Community animation

9.1.1. Leadership, Steering Committees and community service

European XTREEMOS IST-FP6 Integrated Project. Ch. Morin is the *Scientific Coordinator* of the XTREEMOS Integrated Project (http://www.xtreemos.eu/). Y. Jégou is a member of XTREEMOS Executive Committee. Th. Priol is a member of XTREEMOS Scientific Advisory Committee. J.-P. Banâtre is the INRIA representative in the XTREEMOS Governing Board.

European S-CUBE IST-FP7 Network of Excellence J.-L. Pazat is the *Leader* of the Infrastructure JRA 2.3 workpackage of this project. J.-P. Banâtre is the INRIA representative in the S-CUBE Governing Board.

CNRS, GDR ASR. J.-L. Pazat is member of the steering committee of the CNRS Research Co-operative Federation (*Groupement de recherche*, GDR) ASR (*Architectures, Systems and Networks*). F. André serves as the coordinator of the ADAPT action (*Dynamic Adaptation*) of the GSP working group.

ALADDIN-G5K Th. Priol has been the director of ALADDIN-G5K since 2009 and D. Margery is the technical director of ALADDIN-G5K.

9.1.2. Editorial boards, direction of program committees

Ch. Morin served as the Program Chair of CFSE 2009 (*Conférence Francaise sur les systèmes d'exploitation*) held in the framework of Toulouse 2009 (http://www.irit.fr/Toulouse2009/index.html). She was Program Committee Vice-Chair of the IEEE CIT 2009 conference (http://grid.hust.edu.cn/CIT2009). She is the General Chair of the ACM EuroSys 2010 conference (http://eurosys2010.sigops-france.fr/). She has been involved in the Organization Committee of the XTREEMOS summer school held in Oxford, UK in September 2009 (http://www.xtreemos.eu/xtreemos-events/xtreemos-summer-school-2009).

J.-L. Pazat was co-chair of the IEEE CIT2009 program committe, he serves as the Chair of the Organizing Committee of the RenPar, CFSE and Sympa federated conference series. He is the chairman of the Steering Committee of RenPar http://www.renpar.org/).

Th. Priol is a member of the Editorial Board of the *Parallel Computing* Journal and of the *International Journal of Web Services Research*.

P. Linnell has been in charge of the computing infrastructure for the hand-on sessions for the XTREEMOS Summer school, held in Oxford, UK, in September 2009.

9.1.3. Program Committees

Ch. Morin served in the Program Committees of the following conferences: CFSE 2009, HPCVirt 2009, ICA3PP 2009, CIT 2009, GridNets 2009, NAS 2009, SNAPI 2010, Resilience 2010, ISPA 2010, VTC 2010.

J.-L. Pazat served in the Program Committees of the following conferences: CIT 2009, GPC 2009, RenPar 19.

Th. Priol served in the Program Committees of the following conferences: AMCA 2009, CBHPC09, CCGRID09, CLOUD09, ACCS09, GCC09, HPDC09, ICPP09, ICWS09, IWHGA09, SSS09, WGOS09.

Th. Priol served as the PC chair of ParCo09.

9.1.4. Evaluation committees, consulting

Ch. Morin served as a member of the Selection Committee for the INRIA junior researcher competition at INRIA Saclay - Ile de France. She has been appointed by the European Commission as a member of the European Expert Group on Cloud Computing.

Ch. Morin acted as a (primary) referee for the Foreign PhD Committee of Paulo Lopes (Universidade Nova de Lisboa, Portugal).

Th. Priol acts as an expert to review FP6 and FP7 projects for the European Commission as well as ERC proposals.

9.2. Academic teaching

Only the teaching contributions of project-team members on non-teaching positions are mentioned below.

Ch. Morin is responsible for a graduate teaching Module *Distributed Systems: from networks to Grids* of the Master Program in Computer Science, UNIVERSITY RENNES 1. Within this module, she gave lectures on Grid computing. She gave a lecture on cluster single system image operating systems within the *Parallelism* Module of the 3rd-year students of Telecom Sud Paris.

Th. Priol gave lectures on Distributed Shared Memory and Grid Programming within the *Distributed Systems: from Network to Grids* Module of the Master Program, UNIVERSITY RENNES 1.

T. Ropars gave lectures in object oriented programming in Java (INSA, 2nd Year, Main Track) (14h). He also taught lab classes in object oriented programming in Java (INSA, 2nd Year, Main Track) (14h).

9.3. Conferences, seminars, and invitations

Only the events not listed elsewhere are listed below.

Y. Jégou gave a talk entitled *XtreemOS, an Operating System for the Grid* at the NumCoop09 conference, Yaoundé, Cameroun, 2-7 mars 2009. He gave a lecture at the tutorial on Security and Virtual Organization Management in Grids, co-located with the ACM ICS'09 conference, held in New York, USA, in June 2009. He presented demonstrations of the XTREEMOS system at SC'09 (November 2009).

P. Linnell presented demonstrations of the XTREEMOS system at EuroPar 2009 and at SC'09 (November 2009).

J. Gallard gave a talk entitled *VMdeploy: Improving Best-Effort Job Management in Grids* at the SER-OS Workshop, Rennes, France, May 2009. He gave a talk entitled *Discussion: Improving Best-Effort Job Management in Grids* at the Aladdin/G5K Spring school, Nancy, France, April 2009.

Ch. Morin gave a seminar on *XtreemOS Distributed Operating System: from Grids to Clouds*, at ORNL, Oak Ridge, USA, in March 2009. She was invited in a panel on *System Software* at the SOS13 workshop, Hilton Head, USA, in March 2009. She gave a keynote talk on *Virtualization in XtreemOS Distributed Operating System: Requirements & Opportunities* at the HPCVirt 2009 workshop co-located with EuroSys 2009, Nuremberg, Germany, in March 2009. She gave a lecture at the tutorial on Security and Virtual Organization Management in Grids, co-located with the ACM ICS'09 conference, held in New York, USA, in June 2009. She gave a lecture on XTREEMOS at the CEA-EDF-INRIA Computing Science Summer School 2009 on Emerging grid middleware standards, at Saint Lambert des Bois, France, in June 2009. She gave a lecture on *Service Oriented Infrastructures: From Grids to Clouds & XtreemOS Perspective* at XTREEMOS Summer School, Oxford, UK, September 2009. She gave a talk on *Protocols and services for reliable execution of distributed applications in large scale dynamic grids* at the second workshop of the Joint INRIA-NCSA Laboratory for Petascale Computing, Urbana-Champaign, USA, December 2009.

P. Riteau gave a talk entitled *Efficient migration and storage of VMs using distributed content addressing* at the SER-OS 2009 Workshop.

9.4. Administrative responsibilities

T. Priol has been a member of the *Bureau du Comité des projets* until June 2009. From April 2009, he is Scientific Deputy Director at INRIA, responsible for monitoring all research, development, transfer and partnership activities in the thematic areas of Networks and Telecommunication, Distributed and High Performance Computing, Distributed Systems and Services.

P. Riteau is an invited member of the INRIA Rennes - Bretagne Atlantique Center Council (*Comité de Centre*).

9.5. Miscellaneous

F. André has served as a chairwoman for two selection committees of IFSIC (Computer Science department of University of Rennes1), to appoint two assistant Professors in Computer Networks and in Operating Systems.

Ch. Morin has served as a member of the INRIA Rennes local committee for attributing INRIA/CORDIS Ph.D. grants in 2009. She has been a member of the Project-Team Committee of INRIA - Rennes Bretagne Atlantique (*Comité des projets*) since July 2009.

J.-L. Pazat is member of the Steering committee (*c*onseil d'administration) of INSA Rennes. He is a member of the Computer Science Department committee. He is the local coordinator for the international exchange of students at the computer science department of INSA. He is member of the technological development committe (CDT) of IRISA.

T. Priol was a member of the Project-Team Committee of INRIA - Rennes Bretagne Atlantique (*Comité des projets*) until June 2009.

10. Bibliography

Major publications by the team in recent years

[1] M. ALDINUCCI, F. ANDRÉ, J. BUISSON, S. CAMPA, M. COPPOLA, M. DANELUTTO, C. ZOCCOLO. *Parallel program/component adaptivity management*, in "ParCo 2005, Málaga, Spain", 13-16 September 2005.

[2] A. ARENAS, J.-P. BANÂTRE, T. PRIOL. *Developing Secure Chemical Programs with Aspects*, n^o TR-0166, Institute on Knowledge and Data Management , CoreGRID - Network of Excellence, August 2008, http://www.coregrid.net/mambo/images/stories/TechnicalReports/tr-0166.pdf, Technical report.

[3] A. ARENAS, J.-P. BANÂTRE, T. PRIOL. *Developing Autonomic and Secure Virtual Organisations with Chemical Programming*, in "The 11th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS 2009), France Lyon", A. K. DATTA, R. GUERRAOUI, F. PETIT (editors), vol. 5873, Springer, ENS-Lyon, 2009, http://hal.inria.fr/inria-00409502/en/GB.

[4] J.-P. BANÂTRE, D. LE MÉTAYER. *Programming by Multiset Transformation*, in "Communications of the ACM", vol. 36, n^{o} 1, January 1993, p. 98–111.

[5] J. BUISSON, F. ANDRÉ, J.-L. PAZAT. *Supporting Adaptable Applications in Grid Resource Management Systems*, in "8th IEEE/ACM International Conference on Grid Computing, Austin, USA", 19-21 September 2007.

[6] CONSORTIUM, XTREEMOS. *Annex 1 - Description of Work*, April 2006, Contract funded by the European Commission, XtreemOS Integrated Project, IST-033576.

[7] E. JEANVOINE. *Intergiciel pour l´exécution efficace et fiable d´applications distribuées dans des grilles dynamiques de très grande taille*, Université de Rennes 1, IRISA, Rennes, France, November 2007, http://www.irisa.fr/centredoc/publis/theses/irisafolder.2007-03-01.1633154603/irisapublication.2008-01-04.9230494169, In french., Thèse de doctorat.

[8] Z. LAHJOMRI, T. PRIOL. *KOAN: A Shared Virtual Memory for iPSC/2 Hypercube*, in "Proc. of the 2nd Joint Int'l Conf. on Vector and Parallel Processing (CONPAR'92)", Lecture Notes in Computer Science, vol. 634, Springer Verlag, September 1992, p. 441–452, http://hal.inria.fr/inria-00074927.

[9] R. LOTTIAUX. *Gestion globale de la mémoire physique d'une grappe pour un système à image unique : mise en œuvre dans le système Gobelins*, IRISA, Université de Rennes 1, December 2001, Thèse de doctorat.

[10] T. PRIOL. *Efficient support of MPI-based parallel codes within a CORBA-based software infrastructure*, in "Response to the Aggregated Computing RFI from the OMG, Document orbos/99-07-10", July 1999.

[11] Y. RADENAC. *Programmation "chimique" d'ordre supérieur*, Université de Rennes 1, April 2007, Thèse de doctorat.

[12] L. RILLING. *Système d'exploitation à image unique pour une grille de composition dynamique : conception et mise en oeuvre de services fiables pour exécuter les applications distribuées partageant des données*, Université de Rennes 1, IRISA, Rennes, France, November 2005, In French, Thèse de doctorat.

Year Publications

Doctoral Dissertations and Habilitation Theses

[13] B. DAIX. *Abstraction des systèmes informatiques à haute performance pour l'automatisation du déploiement d'applications dynamiques*, Université de Rennes 1, IRISA, Rennes, France, December 2009, In french, Thèse de doctorat.

[14] T. ROPARS. *Services et protocoles pour l'exécution fiable d'applications distribuées dans les grilles de calcul*, Université de Rennes 1, IRISA, Rennes, France, December 2009, In french, Thèse de doctorat.

Articles in International Peer-Reviewed Journal

[15] J.-P. BANÂTRE, T. PRIOL. *Chemical Programming of Future Service-oriented Architectures*, in "Journal of Software (JSW)", vol. 4, 09 2009, p. 738-746, http://hal.inria.fr/inria-00423641/en/.

[16] C. MORIN, Y. JÉGOU, J. GALLARD, P. RITEAU. *Clouds: A New Playground for the XtreemOS Grid Operating System*, in "Parallel Processing Letters (PPL)", vol. 19, n^{o} 3, 2009, p. 435–449, http://hal.inria.fr/inria-00418731/en/.

[17] J.-L. PAZAT, X. SHI. *An adaptive and safe ubicomp for HPC applications*, in "International Journal of Ad Hoc and Ubiquitous Computing", vol. 4, 2009, p. 363–371, http://hal.inria.fr/inria-00435911/en/CN.

International Peer-Reviewed Conference/Proceedings

[18] F. ANDRÉ, G. GAUVRIT, C. PÉREZ. *Dynamic Adaptation of the Master-Worker Paradigm*, in "IEEE 9th International Conference on Computer and Information Technology, Xiamen Chine", 2009, http://hal.archives-ouvertes.fr/inria-00440954/en/.

[19] F. ANDRÉ, M.-T. SEGARRA. *A Distributed Adaptation Model For Ambient Assistive Living Applications*, in "HEALTHINF 2009 - International Conference on Health Informatics, Portugal Porto", IEEE -AAAI, 2009, http://hal.inria.fr/inria-00369379/en/.

[20] F. ANDRÉ, M.-T. SEGARRA. *A Distributed Dynamic Adaptation Model for Component-Based Applications*, in "The 23rd IEEE International Conference on Advanced Information Networking and Applications (AINA2009), Royaume-Uni Bradford", IEEE - University of Bradford, 2009, http://hal.inria.fr/inria-00369323/en/.

[21] F. ANDRÉ, M.-T. SEGARRA. *Building a Context-Aware Ambient Assisted Living Application Using a Self-Adaptive Distributed Model*, in "The Fifth International Conference on Autonomic and Autonomous Systems, Espagne Valencia", IEEE- IARIA, 2009, http://hal.inria.fr/inria-00369327/en/.

[22] F. ANDRÉ, M.-T. SEGARRA, M. ZOUARI. *Distributed Dynamic Self-Adaptation of Data Management in Telemedicine Applications*, in "7th International Conference on Smart Homes and Health Telematics, ICOST 2009, France Tours", 2009, http://hal.inria.fr/inria-00429081/en/.

[23] A. ARENAS, J.-P. BANÂTRE, T. PRIOL. *Developing Autonomic and Secure Virtual Organisations with Chemical Programming*, in "The 11th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS 2009), France Lyon", A. K. DATTA, R. GUERRAOUI, F. PETIT (editors), vol. 5873, Springer, ENS-Lyon, 2009, http://hal.inria.fr/inria-00409502/en/GB.

[24] A. BOUTEILLER, T. ROPARS, G. BOSILCA, C. MORIN, J. DONGARRA. *Reasons for a Pessimistic or Optimistic Message Logging Protocol in MPI Uncoordinated Failure Recovery*, in "IEEE International Conference on Cluster Computing (Cluster 2009), États-Unis d'Amérique New Orleans", 2009, http://hal.inria.fr/inria-00424017/en/US.

[25] J. GALLARD, A. LÈBRE, G. VALLÉE, C. MORIN, P. GALLARD, S. SCOTT. *Refinement Proposal of the Goldberg's Theory*, in "ICA3PP'09: International Conference on Algorithms and Architectures for Parallel Processing, Taiwan Tapei", 2009, p. 853-865, http://hal.inria.fr/inria-00420102/en/US.

[26] J. GALLARD, G. VALLÉE, A. LÈBRE, C. MORIN, P. GALLARD, S. SCOTT. *Complementarity between Virtualization and Single System Image Technologies*, in "Euro-Par 2008 Workshops - Parallel Processing: VHPC 2008, UNICORE 2008, HPPC 2008, SGS 2008, PROPER 2008, ROIA 2008, and DPA 2008, Espagne Las Palmas de Gran Canaria", 2009, http://hal.inria.fr/inria-00419191/en/US.

[27] S. GUELTON, T. GAUTIER, J.-L. PAZAT, S. VARETTE. *Dynamic Adaptation Applied to Sabotage Tolerance*, in "IEEE Conference on Parallel, Distributed and Network-based Processing, Allemagne Weimar", D. E. BAZ, F. SPIES, T. GROSS (editors), Computer Society Press, Euromicro, 2009, http://hal.inria.fr/inria-00369474/en/.

[28] J. MEHNERT-SPAHN, T. ROPARS, M. SCHOETTNER, C. MORIN. *The Architecture of the XtreemOS Grid Checkpointing Service*, in "15th International Euro-Par Conference, Pays-Bas Delft", 2009, http://hal.inria.fr/inria-00424009/en/DE.

[29] P. RITEAU, A. LÈBRE, C. MORIN. *Handling Persistent States in Process Checkpoint/Restart Mechanisms for HPC Systems*, in "9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID '09), China, Shanghai", 2009, http://hal.inria.fr/inria-00424542/en/.

[30] T. ROPARS, C. MORIN. *Active Optimistic Message Logging for Reliable Execution of MPI Applications*, in "15th International Euro-Par Conference, Pays-Bas Delft", 2009, http://hal.inria.fr/inria-00424002/en/.

National Peer-Reviewed Conference/Proceedings

[31] E. DAUBERT, G. GAUVRIT, F. ANDRÉ, O. BARAIS. *Adaptation de services en environnement distribué*, in "RenPar'19, France Toulouse", 2009, http://hal.inria.fr/inria-00411514/en/.

[32] J. GALLARD. *Kget+: An efficient tool for managing VM image snapshots*, in "Toulouse'2009 (RenPar'19 / SympA'13 / CFSE'7) – Poster Session, France Toulouse", 2009, http://hal.inria.fr/inria-00422651/en/.

[33] M. ZOUARI, F. ANDRÉ, M.-T. SEGARRA. *Support d'adaptation dynamique et distribuée dans la conception de systèmes de réplication de données*, in "UbiMob'09, 5èmes journées Francophones Mobilité et Ubiquité, France Lille", 2009, http://hal.inria.fr/inria-00429080/en/.

Research Reports

[34] S. COSTACHE, T. ROPARS, C. MORIN. *Semias: A Framework for Highly Available and Self-Healing Services in Large Scale Dynamic Distributed Systems*, INRIA, 2009, http://hal.inria.fr/inria-00430443/en/, RR-7083, Rapport de recherche.

[35] B. DAIX, S. KORTAS, C. MORIN, C. PÉREZ. *ODD/SAMURAAIE: Providing HPC Users with a Model-Based on-Demand Deployment Architecture*, INRIA, 2009, to appear, Rapport de recherche.

[36] J. GALLARD, A. LÈBRE, C. MORIN. *Saline: Improving Best-Effort Job Management in Grids*, INRIA, 2009, http://hal.inria.fr/inria-00422593/en/, RR-7055, Rapport de recherche.

[37] C. MORIN, J. GALLARD, Y. JÉGOU, P. RITEAU. *Clouds: a New Playground for the XtreemOS Grid Operating System*, INRIA, 2009, http://hal.inria.fr/inria-00358594/en/, RR-6824, Rapport de recherche.

[38] G. VALLÉE, T. NAUGHTON, A. TIKOTEKAR, J. GALLARD, S. SCOTT, C. MORIN. *Architecture for the Next Generation System Management Tools for High Performance Computing Platforms*, INRIA, 2009, http://hal.inria.fr/inria-00424107/en/, RR-7062, Rapport de rechercheUS.

Other Publications

[39] S. JEULAND. *XtreemOS: Dynamic Virtual Organizations*, July 2009, XtreemOS Technical Report #11.

[40] Y. JÉGOU. *Credential Management and Security Architecture in XtreemOS: Proposition for Single-Sign-On and Delegation*, April 2009, XtreemOS Technical Report #7.

[41] Y. JÉGOU. *Supporting interactive jobs in XtreemOS: Proposition*, August 2009, XtreemOS Technical Report #8.

[42] S-CUBE CONSORTIUM. , J.-L. PAZAT (editor)*Basic requirements for Self-healing services and decision support for local adaptation*, March 2009, S-Cube public deliverables.

[43] S-CUBE CONSORTIUM. , Z. NEMETH (editor)*Decision Support for Local Adaptation*, November 2009, S-Cube public deliverables.

[44] S-CUBE CONSORTIUM. , M. PISTORE (editor)*Integration Framework Baseline*, March 2009, S-Cube public deliverables.

[45] S-CUBE CONSORTIUM. , Z. NEMETH (editor)*Use Case Description and State-of-the-Art in Self-* Service Infrastructures and Service Discovery*, April 2009, S-Cube public deliverables.

[46] XTREEMOS CONSORTIUM. , C. MORIN (editor)*Annex 1 - Description of Work (M37-M48)*, June 2009, XtreemOS Integrated Project.

[47] XTREEMOS CONSORTIUM. , C. MORIN (editor)*Annual Project Management Report*, June 2009, XtreemOS Deliverable D1.1.7.

[48] XTREEMOS CONSORTIUM. , C. MORIN (editor)*Collaboration Report and Plan Including Commitment for Contribution to Task 1 to 8*, September 2009, XtreemOS Deliverable D5.3.4.

[49] XTREEMOS CONSORTIUM. *Communication and Dissemination Report and Plan*, June 2009, XtreemOS Deliverable D5.1.8.

[50] XTREEMOS CONSORTIUM. *Evaluation Report*, December 2009, XtreemOS Deliverable D4.2.6.

[51] XTREEMOS CONSORTIUM. , C. MORIN (editor)*First Advanced Release of Linux-XOS*, June 2009, XtreemOS Deliverable D2.1.7.

[52] XTREEMOS CONSORTIUM. , C. MORIN (editor)*Prototype of the First Advanced Version of LinuxSSI*, June 2009, XtreemOS Deliverable D2.2.10.

[53] XTREEMOS CONSORTIUM. *Public LiveCD release*, December 2009, XtreemOS Deliverable D4.1.30.

[54] XTREEMOS CONSORTIUM. *Public Release of XtreemOS 2.0*, December 2009, XtreemOS Deliverable D4.1.23.

[55] XTREEMOS CONSORTIUM. *Training Report and Plan*, May 2009, XtreemOS Deliverable D5.2.3.

[56] XTREEMOS CONSORTIUM. , M. COPPOLA (editor)*XtreemOS User's Guide*, October 2009, http://www.xtreemos.eu/software/userguide.pdf, XtreemOS Technical Report #10.

[57] XTREEMOS CONSORTIUM. , M. COPPOLA (editor)*XtreemOS, Advanced Guide: Installation and Administration*, October 2009, http://www.xtreemos.eu/software/adminguide.pdf, XtreemOS Technical Report #9.

References in notes

[58] *Dynaco Web site*, http://dynaco.gforge.inria.fr/.

[59] I. FOSTER, C. KESSELMAN (editors). *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann Publishers, 1998.

[60] *Wireless Application Protocol 2.0: technical white paper*, January 2002, http://www.wapforum.org/what/WAPWhite_Paper1.pdf.

[61] R. ARMSTRONG, D. GANNON, A. GEIST, K. KEAHEY, S. KOHN, L. MCINNES, S. PARKER, B. SMOLINSKI. *Toward a Common Component Architecture for High-Performance Scientific Computing*, in "Proceeding of the 8th IEEE International Symposium on High Performance Distributed Computation", August 1999.

[62] J.-P. BANÂTRE, P. FRADET, Y. RADENAC. *Principles of Chemical Programming*, in "Proceedings of the 5th International Workshop on Rule-Based Programming (RULE 2004)", S. ABDENNADHER, C. RINGEISSEN (editors), ENTCS, vol. 124, n° 1, Elsevier, June 2005, p. 133–147.

[63] J.-P. BANÂTRE, P. FRADET, D. LE MÉTAYER. *Gamma and the Chemical Reaction Model: Fifteen Years After*, in "Multiset Processing", LNCS, vol. 2235, Springer-Verlag, 2001, p. 17–44.

[64] J.-P. BANÂTRE, D. LE MÉTAYER. *A new computational model and its discipline of programming*, n° RR0566, INRIA, September 1986, http://hal.inria.fr/inria-00075988, Technical report.

[65] J.-P. BANÂTRE, D. LE MÉTAYER. *Programming by Multiset Transformation*, in "Communications of the ACM", vol. 36, n° 1, January 1993, p. 98–111.

[66] F. BAUDE, D. CAROMEL, C. DALMASSO, M. DANELUTTO, V. GETOV, L. HENRIO, C. PÉREZ. *GCM: A Grid Extension to Fractal for Autonomous Distributed Components*, in "Annals of Telecommunications", 2008.

[67] G. BERRY, G. BOUDOL. *The Chemical Abstract Machine*, in "Theoretical Computer Science", vol. 96, 1992, p. 217–248.

[68] E. BRUNETON, T. COUPAYE, M. LECLERCQ, V. QUÉMA, J.-B. STEFANI. *The FRACTAL component model and its support in Java: Experiences with Auto-adaptive and Reconfigurable Systems*, in "Softw. Pract. Exper.", vol. 36, n° 11-12, 2006, p. 1257–1284.

[69] J. BUISSON. *Adaptation dynamique de programmes et composants parallèles*, INSA de Rennes, 25 September 2006, Ph. D. Thesis.

[70] D. CHEFROUR, F. ANDRÉ. *Auto-adaptation de composants ACEEL coopérants*, in "3e Conférence française sur les systèmes d'exploitation (CFSE 3)", 2003.

[71] J. GALLARD, A. LÈBRE, C. MORIN. *Saline: Improving Best-Effort Job Management in Grids*, in "PDP 2010: The 18th Euromicro International Conference on Parallel, Distributed and Network-Based Computing – Special Session: Virtualization, Pisa, Italy", 2010, to appear.

[72] K. GHARACHORLOO, D. LENOSKI, J. LAUDON, P. GIBBONS, A. GUPTA, J. HENESSY. *Memory Consistency and event ordering in scalable shared memory multiprocessors*, in "17th Annual Intl. Symposium on Computer Architectures (ISCA)", ACM, May 1990, p. 15–26.

[73] J. GRAY, D. SIEWIOREK. *High Availability Computer Systems*, in "IEEE Computer", September 1991.

[74] E. JEANNOT, B. KNUTSSON, M. BJORKMANN. *Adaptive Online Data Compression*, in "IEEE High Performance Distributed Computing (HPDC 11)", 2002.

[75] P. KELEHER, A. COX, W. ZWAENEPOEL. *Lazy Release Consistency for Software Distributed Shared Memory*, in "19th Intl. Symposium on Computer Architecture", May 1992, p. 13–21.

[76] P. KELEHER, D. DWARKADAS, A. COX, W. ZWAENEPOEL. *TreadMarks: Distributed Shared Memory on standard workstations and operating systems*, in "Proc. 1994 Winter Usenix Conference", January 1994, p. 115–131.

[77] L. LAMPORT. *How to Make a Multiprocessor Computer That Correctly Executes Multiprocess Programs*, in "IEEE Transactions on Computers", vol. 28, n° 9, September 1979, p. 690–691.

[78] P. LEE, T. ANDERSON. *Fault Tolerance: Principles and Practice*, vol. 3 of Dependable Computing and Fault-Tolerant Systems, Springer Verlag, second revised edition, 1990.

[79] A. LÈBRE, R. LOTTIAUX, E. FOCHT, C. MORIN. *Reducing Kernel Development Complexity in Distributed Environments*, in "Euro-Par", 2008, p. 576-586.

[80] F. MATTERN. *Virtual Time and Global States in Distributed Systems*, in "Proc. Int. Workshop on Parallel and Distributed Algorithms, Gers, France", North-Holland, 1989, p. 215–226.

[81] D. S. MILOJICIC, V. KALOGERAKI, R. LUKOSE, K. NAGARAJA, J. PRUYNE, B. RICHARD, S. ROLLINS, Z. XU. *Peer-to-Peer Computing*, n° HPL-2002-57, HP Labs, March 2002, http://www.hpl.hp.com/techreports/2002/HPL-2002-57R1.pdf, Submitted to Computing Surveys, Research Report.

[82] OSGI ALLIANCE. *OSGi Service Platform Core Specification, Release 4 Version 4.2*, September 2007, http://www.osgi.org.

[83] M. P. PAPAZOGLOU, P. TRAVERSO, S. DUSTDAR, F. LEYMANN. *Service-Oriented Computing: State of the Art and Research Challenges*, in "Computer", vol. 40, 2007, p. 38-45.

[84] G. PĂUN. *Computing with Membranes*, in "Journal of Computer and System Sciences", vol. 61, n^{o} 1, 2000, p. 108-143.

[85] C. SZYPERSKI. *Component Software - Beyond Object-Oriented Programming*, Addison-Wesley / ACM Press, 1998.