IN PARTNERSHIP WITH:
CNRS

Institut polytechnique de Grenoble

Université Joseph Fourier (Grenoble)

Activity Report 2015

Project-Team MESCAL

Middleware efficiently scalable

IN COLLABORATION WITH: Laboratoire d'Informatique de Grenoble (LIG)

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Distributed and High Performance Computing

Table of contents

Project-Team MESCAL

Creation of the Project-Team: 2006 January 01, end of the Project-Team: 2015 December 31

Keywords:

Computer Science and Digital Science:

1. - Architectures, systems and networks
1.1.4. - High performance computing
1.1.6. - Cloud
1.2.3. - Routing
1.2.4. - QoS, performance evaluation
2.6. - Infrastructure software
2.6.1. - Operating systems
2.6.2. - Middleware
2.6.3. - Virtual machines
3.4.4. - Optimization and learning
6. - Modeling, simulation and control
6.1.4. - Multiscale modeling
6.2.3. - Probabilistic methods
6.2.4. - Statistical methods
6.2.6. - Optimization
6.2.7. - High performance computing
6.4.2. - Stochastic control
7.1. - Parallel and distributed algorithms
7.11. - Performance evaluation
7.3. - Operations research, optimization, game theory

Other Research Topics and Application Domains:

4.3.1. - Smart grids
4.4.1. - Green computing
6.2.1. - Wired technologies
6.2.2. - Radio technology
6.3.2. - Network protocols
7. - Transport and logistics
9.6. - Reproducibility

1. Members

Research Scientists

Bruno Gaujal [Team leader, Inria, Senior Researcher, HdR]
Nicolas Gast [Inria, Researcher]
Arnaud Legrand [CNRS, Researcher, HdR]
Panayotis Mertikopoulos [CNRS, Researcher]

Faculty Members

Elena-Veronica Belmega [ENSEA, Associate Professor, from Sep 2015]

Yves Denneulin [INP Grenoble, Professor, HdR]
Florence Perronnin [Univ. Grenoble I, Associate Professor]
Brigitte Plateau [INP Grenoble, Professor, HdR]
Olivier Richard [Univ. Grenoble I, Associate Professor]
Jean-Marc Vincent [Univ. Grenoble I, Associate Professor]

Engineers

Maxime Boutserin [Inria, until Sep 2015]
Benjamin Briot [Inria]
Romain Cavagna [Univ. Grenoble I]
Youenn Corre [Inria]
Augustin Degomme [CNRS, until Jan 2015]
Pierre Neyron [CNRS]
Generoso Pagano [Inria, until Jun 2015, granted by OSEO Innovation]
Baptiste Pichot [Inria, from Oct 2015]
Christian Seguy [CNRS]
Bruno Bzeznik [Univ. Grenoble I]

PhD Students

Joaquim Carvalho Assuncao [PUC RS, From Jan 2015 to Dec 2015]
Stéphane Durand [Univ. Grenoble I, From Oct 2015]
Vinicius Garcia Pinto [UFRGS/Univ. Grenoble I, From Sep 2015]
Franz Christian Heinrich [Univ. Grenoble I, Inria grant, from Dec 2015]
Rafael Keller Tesser [UFRGS, from August 2015]
Alexis Martin [Univ Grenoble I, Inria, granted by OSEO Innovation]
Erick Ramon Meneses Cuadros [Univ. Grenoble 1, CIFRE Orange, until Jun 2015]
Michael Mercier [Univ. Grenoble I, CIFRE ATOS, starting in October 2015]
Benoît Vinot [Univ. Grenoble I, from May 2015, granted by CIFRE Schneider]

Post-Doctoral Fellows

Josu Doncel [Inria]
Guillaume Massonnet [Inria]
Luka Stanisic [Inria]
Brice Videau [CNRS]
Angelika Studeny [Univ. Paris VII, until Nov 2015]

Visiting Scientist

Lucas Mello Schnorr [Prof. UFRGS, Sep 2015]

Administrative Assistant

Annie Simon [Inria]

Others

Mathieu Baille [Inria, Intern L3, Jun 2015 until Aug 2015]
Amaury Bouchra Pilet [Inria, Intern L3, from Jun 2015 until Jul 2015]
Marjan Bozorg [CNRS, Intern M2, from Mar 2015 until Jul 2015]
Steven Quinito Masnada [CNRS, Intern M1, from Feb 2015 until Jun 2015]
Thibaud Buchs [CNRS, Intern M1, from Feb 2015 until Aug 2015]
Jing Han [Inria, Intern M1, from Feb 2015 until Jun 2015]
Baptiste Jonglez [ENS Lyon, Intern ENS-4A, from Sep 2015]
Florian Popek [Inria, Intern L3, Jul 2015]
Wei Wei [Inria, Intern M1, from Jun 2015 until Jul 2015]
Xin Zhao [Inria, Intern M1, from Feb 2015 until Jul 2015]

2. Overall Objectives

2.1. Presentation

MESCAL is a project-team of Inria jointly with UJF and Grenoble INP universities and CNRS, created in 2006 as an offspring of the former APACHE project-team, together with MOAIS.

MESCAL's research activities and objectives were evaluated by Inria in 2012. The MESCAL project-team received positive evaluations and useful feedback. The project-team was extended for another 4 years by the Inria evaluation commission.

2.2. Objectives

The recent evolutions in network and computer technology, as well as their diversification, go with a tremendous change in the use of these architectures: applications and systems can now be designed at a much larger scale than before. This scaling evolution concerns at the same time the amount of data, the number and heterogeneity of processors, the number of users, and the geographical diversity of the users.

This race towards *large scale* questions many assumptions underlying parallel and distributed algorithms as well as operating middleware. Today, most software tools developed for average size systems cannot be run on large scale systems without a significant degradation of their performances.

The goal of the MESCAL project-team is to design and validate efficient exploitation mechanisms (algorithms, middleware and system services) for large distributed infrastructures.

MESCAL's target infrastructures are grids obtained through sharing of available resources inside autonomous computing services, lightweight grids (such as the local CIMENT Grid), clusters of intranet resources (Condor) or aggregation of Internet resources (SETI@home, BOINC) as well as clouds (Amazon, Google clouds) and communication networks (5G, LTE and Wifi networks).

Application domains concern intensive scientific computations and low power high performance computing. We are also designing algorithms and middleware for SON (Self Organizing Networks) with implementations in wireless devices and base stations. Our range of applications also include the power grid (smart grids) as well as shared transportation systems.

MESCAL's methodology in order to ensure efficiency and scalability of proposed mechanisms is based on mathematical modeling and performance evaluation of the full spectrum of large scale systems from target architectures, software layers to applications.

3. Research Program

3.1. Large System Modeling and Analysis

Participants: Nicolas Gast, Bruno Gaujal, Arnaud Legrand, Panayotis Mertikopoulos, Florence Perronnin, Olivier Richard, Jean-Marc Vincent.

Markov chains, Queuing networks, Mean field approximation, Simulation, Performance evaluation, Discrete event dynamic systems.

3.1.1. Simulation of distributed systems

Since the advent of distributed computer systems, an active field of research has been the investigation of *scheduling* strategies for parallel applications. The common approach is to employ scheduling heuristics that approximate an optimal schedule. Unfortunately, it is often impossible to obtain analytical results to compare the efficiency of these heuristics. One possibility is to conduct large numbers of back-to-back experiments on real platforms. While this is possible on tightly-coupled platforms, it is unfeasible on modern distributed platforms (i.e., grids or peer-to-peer environments) as it is labor-intensive and does not enable repeatable results. The solution is to resort to *simulations*.

3.1.1.1. Flow Simulations

To make simulations of large systems efficient and trustful, we have used flow simulations (where streams of packets are abstracted into flows). SimGrid is a simulation platform that specifically targets the simulation of large distributed systems (grids, clusters, peer-to-peer systems, volunteer computing systems, clouds) from the perspective of applications. It enables to obtain repeatable results and to explore wide ranges of platform and application scenarios.

3.1.1.2. Perfect Simulation

Using a constructive representation of a Markovian queuing network based on events (often called GSMPs), we have designed perfect simulation algorithms computing samples distributed according to the stationary distribution of the Markov process with no bias. The tools based on our algorithms (ψ) can sample the stationary measure of Markov processes using directly the queuing network description. Some monotone networks with up to 10^{50} states can be handled within minutes over a regular PC.

3.1.2. Fluid models and mean field limits

When the size of systems grows very large, one may use asymptotic techniques to get a faithful estimate of their behavior. One such tool is mean field analysis and fluid limits, that can be used at a modeling and simulation level. Proving that large discrete dynamic systems can be approximated by continuous dynamics uses the theory of stochastic approximation pioneered by Michel Benaïm or population dynamics introduced by Thomas Kurtz and others. We have extended the stochastic approximation approach to take into account discontinuities in the dynamics as well as to tackle optimization issues.

Recent applications include call centers and peer to peer systems, where the mean field approach helps to get a better understanding of the behavior of the system and to solve several optimization problems. Another application concerns task brokering in desktop grids taking into account statistical features of tasks as well as of the availability of the processors. Mean field has also been applied to the performance evaluation of work stealing in large systems and to model central/local controllers as well as knitting systems.

3.1.3. Game Theory

Resources in large-scale distributed platforms (grid computing platforms, enterprise networks, peer-to-peer systems) are shared by a number of users having conflicting interests who are thus prone to act selfishly. A natural framework for studying such non-cooperative individual decision-making is game theory. In particular, game theory models the decentralized nature of decision-making.

It is well known that such non-cooperative behaviors can lead to important inefficiencies and unfairness. In other words, individual optimizations often result in global resource waste. In the context of game theory, a situation in which all users selfishly optimize their own utility is known as a *Nash equilibrium* or *Wardrop equilibrium*. In such equilibria, no user has interest in unilaterally deviating from its strategy. Such policies are thus very natural to seek in fully distributed systems and have some stability properties. However, a possible consequence is the *Braess paradox* in which the increase of resource happens at the expense of *every* user. This is why, the study of the occurrence and degree of such inefficiency is of crucial interest. Up until now, little is known about general conditions for optimality or degree of efficiency of these equilibria, in a general setting.

Many techniques have been developed to enforce some form of collaboration and improve these equilibria. In this context, it is generally prohibitive to take joint decisions so that a global optimization cannot be achieved. A possible option relies on the establishment of virtual prices, also called *shadow prices* in congestion networks. These prices ensure a rational use of resources.

Once the payoffs are fixed (using shadow prices or not), the main question is to design algorithms that allow the players to learn Nash equilibria in a distributed way, while being robust to noise and information delay as well as fast enough to outrate changing conditions of the environment.

3.2. Management of Large Architectures

Participants: Nicolas Gast, Arnaud Legrand, Olivier Richard.

Administration, Deployment, Peer-to-peer, Clusters, Grids, Clouds, Job scheduler

3.2.1. Instrumentation, analysis and prediction tools

To understand complex distributed systems, one has to provide reliable measurements together with accurate models before applying this understanding to improve system design.

Our approach for instrumentation of distributed systems (embedded systems as well as multi-core machines or distributed systems) relies on quality of service criteria. In particular, we focus on non-obtrusiveness and experimental reproducibility.

Our approach for analysis is to use statistical methods with experimental data of real systems to understand their normal or abnormal behavior. With that approach we are able to predict availability of very large systems (with more than 100,000 nodes), to design cost-aware resource management (based on mathematical modeling and performance evaluation of target architectures), and to propose several scheduling policies tailored for unreliable and shared resources.

3.2.2. Fairness in large-scale distributed systems

Large-scale distributed platforms (grid computing platforms, enterprise networks, peer-to-peer systems) result from the collaboration of many people. Thus, the scaling evolution we are facing is not only dealing with the amount of data and the number of computers but also with the number of users and the diversity of their behavior. In a high-performance computing framework, the rationale behind this joining of forces is that most users need a larger amount of resources than what they have on their own. Some only need these resources for a limited amount of time. On the opposite some others need as many resources as possible but do not have particular deadlines. Some may have mainly tightly-coupled applications while some others may have mostly embarrassingly parallel applications. The variety of user profiles makes resources sharing a challenge. However resources have to be *fairly* shared between users, otherwise users will leave the group and join another one. Large-scale systems therefore have a real need for fairness and this notion is missing from classical scheduling models.

3.2.3. Tools to operate clusters

The MESCAL project-team studies and develops a set of tools designed to help the installation and the use of a cluster of PCs. The first version had been developed for the Icluster1 platform exploitation. The main tools are a scalable tool for cloning nodes (KA-DEPLOY) and a parallel launcher based on the TAKTUK project (now developed by the MOAIS project-team). Many interesting issues have been raised by the use of the first versions among which we can mention environment deployment, robustness and batch scheduler integration. A second generation of these tools is thus under development to meet these requirements.

KA-DEPLOY has been retained as the primary deployment tool for the experimental national grid Grid'5000.

3.2.4. Simple and scalable batch scheduler for clusters and grids

Most known batch schedulers (PBS, LSF, Condor, ...) are built in a monolithic way, with the purpose of fulfilling most of the exploitation needs. This results in systems of high software complexity (150,000 lines of code for OpenPBS), offering a growing number of functions that are, most of the time, not used. In such a context, it becomes hard to control both the robustness and the scalability of the whole system.

OAR is an attempt to address these issues. Firstly, OAR is written in a very high level language (Perl) and makes intensive use of high level tools (MySql and TAKTUK), thereby resulting in a concise code (around 5000 lines of code) easy to maintain and extend. This small code as well as the choice of widespread tools (MySql) are essential elements that ensure a strong robustness of the system. Secondly, OAR makes use of SQL queries to perform most of its job management tasks thereby getting advantage of the strong scalability of most database management tools. Such scalability is further improved in OAR by making use of TAKTUK to manage nodes themselves.

3.3. Migration resilience; Large scale data management

Participant: Yves Denneulin.

Fault tolerance, migration, distributed algorithms.

Most propositions to improve reliability address only a given application or service. This may be due to the fact that until clusters and intranet architectures arose, it was obvious that client and server nodes were independent. This is not the case in parallel scientific computing where a fault on a node can lead to a data loss on thousands of other nodes. MESCAL's work on this topic is based on the idea that each process in a parallel application will be executed by a group of nodes instead of a single node: when the node in charge of a process fails, another in the same group can replace it in a transparent way for the application.

There are two main problems to be solved in order to achieve this objective. The first one is the ability to migrate processes of a parallel, and thus communicating, application without enforcing modifications. The second one is the ability to maintain a group structure in a completely distributed way. They both rely on a close interaction with the underlying operating systems and networks, since processes can be migrated in the middle of a communication. This can only be done by knowing how to save and replay later all ongoing communications, independently of the communication pattern. Freezing a process to restore it on another node is also an operation that requires collaboration of the operating system and a good knowledge of its internals. The other main problem (keeping a group structure) belongs to the distributed algorithms domain and is of a higher level nature.

4. Application Domains

4.1. Cloud, Grid, Multi-core and Desktop Computing

Participants: Arnaud Legrand, Olivier Richard, Jean-Marc Vincent.

Software tools were developed to carry experiments on clouds and grids (Kameleon and Expo). Other tools (Pajé, Viva, Framesoc and Ocelotl) have been designed to monitor, trace and analyse applications running on multi-core and grid computers Such traces have also been used in SIMGRID to simulate volunteer computing systems at unprecedented scale.

4.2. Wireless Networks

Participants: Bruno Gaujal, Panayotis Mertikopoulos.

MESCAL is involved in the common laboratory between Inria and Alcatel-Lucent. Bruno Gaujal is leading the Selfnets research action. This action was started in 2008 and was renewed for four more years (from 2012 to 2016). In our collaboration with Alcatel we use game theory techniques as well as evolutionary algorithms to compute optimal configurations in wireless networks (typically 3G or LTE networks) in a distributed manner. We have also been working on optimal spectrum management of MIMO systems, routing in ad-hoc works and power allocation in future 5G networks.

4.3. On-demand Geographical Maps

Participant: Jean-Marc Vincent.

This joint work involves the UMR 8504 Géographie-Cité, LIG, UMS RIATE and the Maisons de l'Homme et de la Société.

Improvements in the Web developments have opened new perspectives in interactive cartography. Nevertheless existing architectures have some problems to perform spatial analysis methods that require complex calculus over large data sets. Such a situation involves some limitations in the query capabilities and analysis methods proposed to users. The HyperCarte consortium with LIG, Géographie-cité and UMR RIATE proposes innovative solutions to these problems. Our approach deals with various areas such as spatio-temporal modeling, parallel computing and cartographic visualization that are related to spatial organizations of social phenomena.

4.4. Energy and Transportation

Participant: Nicolas Gast.

This work is mainly done within the Quanticol European project.

Smart urban transport systems and smart grids are two examples of collective adaptive systems. They consist of a large number of heterogeneous entities with decentralised control and varying degrees of complex autonomous behaviour. Within the QUANTICOL project, we develop an analysis tools to help to reason about such systems. Our work relies on tools from fluid and mean-field approximation to build decentralized algorithms that solve complex optimization problems. We focus on two problems: decentralized control of electric grids and capacity planning in vehicle-sharing systems to improve load balancing.

5. New Software and Platforms

5.1. CiGri

FUNCTIONAL DESCRIPTION

CiGri is a middleware which gathers the unused computing resource from intranet infrastructure and makes it available for the processing of large set of tasks. It manages the execution of large sets of parametric tasks on lightweight grid by submitting individual jobs to each batch scheduler. It is s associated to the OAR resource management system (batch scheduler). Users can easily monitor and control their set of jobs through a web portal. CiGri provides mechanisms to identify job error causes, to isolate faulty components and to resubmit jobs in a safer context.

- Contact: Olivier Richard
- URL: https://www.projet-plume.org/fiche/cigri

5.2. ComputeMode

ComputeMode: On-demand HPC cluster manager
KEYWORDS: HPC - Clusters - Operating system provisioning
FUNCTIONAL DESCRIPTION

ComputeMode is a on-demand HPC cluster manager, it allows deploying lightweight clustering framework on intranets.

ComputeMode is a software infrastructure that allows to extend or create a Grid through the aggregation of unused computing resources. For instance, a virtual cluster can be built using anyone's PC while not in use. Indeed, most PCs in large companies or university campus are idle at night, on weekends, and during vacations, training periods or business trips.

The main benefits of ComputeMode are the following

Easy deployment: the integration into an existing infrastructure is very easy: no modification is required on your PCs. ComputeMode comes as a software-only solution. The integration with major batch manager systems such as Sun Grid Engine, Platform LSF and Portable Batch System (PBS) can also be achieved. Seamless integration for the scientist: he/she submits unmodified computational jobs through his/her usual interface (batch submission engine), just like with any Beowulf type cluster. Seamless integration for the PC owner/user: ComputeMode runs when his/her PC is idle (night, weekends, ...) so annoyance is minimal if existant

Using ComputeMode, the life cycle of the PCs is basically split between 2 modes of operation a user mode, where the company's installation of Microsoft Windows or GNU/Linux remains a computation mode (hence the product name): uses a diskless boot of a GNU/Linux system and offers the PC's CPU power, RAM and connectivity to the Grid.

- Participants: Pierre Neyron, Olivier Richard and Bruno Bzeznik
- Partners: LIG - ANDRA
- Contact: Olivier Richard
- URL: http://computemode.imag.fr

5.3. Framesoc

FUNCTIONAL DESCRIPTION

Framesoc is the core software infrastructure of the SoC-Trace project. It provides a graphical user environment for execution-trace analysis, featuring interactive analysis views as Gantt charts or statistics views. It provides also a software library to store generic trace data, play with them, and build other analysis tools (e.g., Ocelotl).

- Participants: Jean-Marc Vincent and Arnaud Legrand
- Contact: Jean-Marc Vincent
- URL: http://soctrace-inria.github.io/framesoc/

5.4. GameSeer

FUNCTIONAL DESCRIPTION

GameSeer is a tool for students and researchers in game theory that uses Mathematica to generate phase portraits for normal form games under a variety of (user-customizable) evolutionary dynamics. The whole point behind GameSeer is to provide a dynamic graphical interface that allows the user to employ Mathematica's vast numerical capabilities from a simple and intuitive front-end. So, even if you've never used Mathematica before, you should be able to generate fully editable and customizable portraits quickly and painlessly.

- Contact: Panayotis Mertikopoulos
- URL: http://mescal.imag.fr/membres/panayotis.mertikopoulos/publications.html

5.5. KA-Tools

FUNCTIONAL DESCRIPTION

The KA-Tools is a software suite developed by MESCAL for exploitation of clusters and grids. It uses a parallelization technique based on spanning trees with a recursive starting of programs on nodes. Industrial collaborations were carried out with Mandrake, BULL, HP and Microsoft.

- Contact: Olivier Richard
- URL: http://ka-tools.imag.fr/

5.6. Kadeploy

KEYWORD: Operating system provisioning
FUNCTIONAL DESCRIPTION

Kadeploy is a scalable, efficient and reliable deployment system (cluster provisioning solution) for cluster and grid computing. It provides a set of tools for cloning, configuring (post installation) and managing cluster nodes. It can deploy a 300-nodes cluster in a few minutes, without intervention from the system administrator.

- Participants: Emmanuel Jeanvoine, Olivier Richard, Lucas Nussbaum and Luc Sarzyniec
- Partners: CNRS - Université de Lorraine - Loria - Grid'5000 - Inria
- Contact: Olivier Richard
- URL: http://kadeploy3.gforge.inria.fr

5.7. Kameleon

FUNCTIONAL DESCRIPTION

Kameleon is a simple but powerful tool to generate customized appliances. With Kameleon, you make your recipe that describes how to create step by step your own distribution. At start Kameleon is used to create custom kvm, docker, VirtualBox, ..., but as it is designed to be very generic you can probably do a lot more than that.

- Participant: Olivier Richard
- Partner: Grid'5000
- Contact: Olivier Richard
- URL: http://kameleon.imag.fr/

5.8. OAR

KEYWORDS: HPC - Cloud - Clusters - Resource manager - Light grid
SCIENTIFIC DESCRIPTION

This batch system is based on a database (PostgreSQL (preferred) or MySQL), a script language (Perl) and an optional scalable administrative tool (e.g. Taktuk). It is composed of modules which interact mainly via the database and are executed as independent programs. Therefore, formally, there is no API, the system interaction is completely defined by the database schema. This approach eases the development of specific modules. Indeed, each module (such as schedulers) may be developed in any language having a database access library.
FUNCTIONAL DESCRIPTION

OAR is a versatile resource and task manager (also called a batch scheduler) for HPC clusters, and other computing infrastructures (like distributed computing experimental testbeds where versatility is a key).

- Participants: Olivier Richard, Pierre Neyron, Salem Harrache and Bruno Bzeznik
- Partners: LIG - CNRS - Grid'5000 - CIMENT
- Contact: Olivier Richard
- URL: http://oar.imag.fr

5.9. Ocelotl

Multidimensional Overviews for Huge Trace Analysis
FUNCTIONAL DESCRIPTION

Ocelotl is an innovative visualization tool, which provides overviews for execution trace analysis by using a data aggregation technique. This technique enables to find anomalies in huge traces containing up to several billions of events, while keeping a fast computation time and providing a simple representation that does not overload the user.

- Participants: Arnaud Legrand and Jean-Marc Vincent
- Contact: Jean-Marc Vincent
- URL: http://soctrace-inria.github.io/ocelotl/

5.10. PEPS

FUNCTIONAL DESCRIPTION

The main objective of PEPS is to facilitate the solution of large discrete event systems, in situations where classical methods fail. PEPS may be applied to the modelling of computer systems, telecommunication systems, road traffic, or manufacturing systems.

- Participants: Luka Stanisic, Arnaud Legrand, Augustin Degomme, Jean-Marc Vincent and Florence Perronnin
- Contact: Arnaud Legrand
- URL: http://www-id.imag.fr/Logiciels/peps/

5.11. PSI

Perfect Simulator
FUNCTIONAL DESCRIPTION

Perfect simulator is a simulation software of markovian models. It is able to simulate discrete and continuous time models to provide a perfect sampling of the stationary distribution or directly a sampling of functional of this distribution by using coupling from the past. The simulation kernel is based on the CFTP algorithm, and the internal simulation of transitions on the Aliasing method.

- Contact: Arnaud Legrand
- URL: https://gforge.inria.fr/projects/psi/

5.12. Pajé

FUNCTIONAL DESCRIPTION

The Pajé generic tool provides interactive and scalable behavioral visualizations of parallel and distributed applications, helping to capture the dynamics of their executions, because of its genericity, it can be used unchanged in a large variety of contexts.

- Participants: Arnaud Legrand and Jean-Marc Vincent
- Contact: Jean-Marc Vincent
- URL: http://paje.sourceforge.net/

5.13. PajéNG

Pajé Next Generation
FUNCTIONAL DESCRIPTION

Pajé Next Generation is a re-implementation (in C++) and direct heir of the well-known Paje visualization tool for the analysis of execution traces (in the Paje File Format) through trace visualization (space/time view). The tool is released under the GNU General Public License 3. PajeNG comprises the libpaje library, the space-time visualization tool in pajeng and a set of auxiliary tools to manage Paje trace files (such as pj_dump and pj_validate).

- Participants: Jean-Marc Vincent and Arnaud Legrand
- Contact: Jean-Marc Vincent
- URL: https://github.com/schnorr/pajeng

5.14. SimGrid

KEYWORDS: Large-scale Emulators - Grid Computing - Distributed Applications
FUNCTIONAL DESCRIPTION

Scientific Instrument for the study of Large-Scale Distributed Systems. SimGrid is a toolkit that provides core functionalities for the simulation of distributed applications in heterogeneous distributed environments.

- Participants: Jonathan Rouzaud-Cornabas, Frédéric Suter, Martin Quinson, Arnaud Legrand, Takahiro Hirofuchi, Adrien Lèbre, Jonathan Pastor, Mario Südholt, Flavien Quesnel, Luka Stanisic, Augustin Degomme, Jean-Marc Vincent and Florence Perronnin
- Partners: CNRS - Université de Nancy - University of Hawaii - Université de Reims Champagne-Ardenne - Femto-st
- Contact: Arnaud Legrand
- URL: http://simgrid.gforge.inria.fr/

5.15. Viva

FUNCTIONAL DESCRIPTION

Viva is an open-source tool used to analyze traces (in the Paje File Format) registered during the execution of parallel or distributed applications. The tool also serves as a sandbox to the development of new visualization techniques. Current features include: Temporal integration using dynamic time-intervals Spatial aggregation through hierarchical traces Interactive Graph Visualization with a force-directed algorithm, with viva Squarified Treemap to compare processes behavior on scale, with vv_treemap.

- Contact: Arnaud Legrand
- URL: https://github.com/schnorr/viva

5.16. Platforms

5.16.1. Grid'5000

The MESCAL project-team is involved in development and management of Grid'5000 platform. The Digitalis and IDPot clusters are integrated in Grid'5000 as well as of CIMENT.

5.16.2. Local cluster computing platforms: ICluster-2, IDPot, Digitalis

The MESCAL project-team manages a cluster computing center on the Grenoble campus. The center manages different architectures: a 48 bi-processors PC (ID-POT), and the center is involved with a cluster based on 110 bi-processors Itanium2 (ICluster-2) and another based on 34 bi-processor quad-core XEON (Digitalis) located at Inria. The three of them are integrated in the Grid'5000 grid platform.

More than 60 research projects in France have used the architectures, especially the 204 processors Icluster-2. Half of them have run typical numerical applications on this machine, the remainder has worked on middleware and new technology for cluster and grid computing. The Digitalis cluster is also meant to replace the Grimage platform in which the MOAIS project-team is very involved.

5.16.3. The Bull Machine

In the context of our collaboration with Bull the MESCAL project-team exploits a Novascale NUMA machine. The configuration is based on 8 Itanium II processors at 1.5 Ghz and 16 GB of RAM. This platform is mainly used by the Bull PhD students. This machine is also connected to the CIMENT Grid.

6. New Results

6.1. Reproducible Research

In the field of large-scale distributed systems, experimentation is particularly difficult. The studied systems are complex, often non-deterministic and unreliable, software is plagued with bugs, whereas the experiment workflows are unclear and hard to reproduce. In [11], we provide an extensive list of features offered by general-purpose experiment management tools dedicated to distributed systems research on real platforms. We then use it to assess existing solutions and compare them, outlining possible future paths for improvements.

In [20], we address the question of developing a lightweight and effective workflow for conducting experimental research on modern parallel computer systems in a reproducible way. Our workflow simply builds on two well-known tools (Org-mode and Git) and enables us to address issues such as provenance tracking, experimental setup reconstruction, replicable analysis. Although this workflow is perfectible and cannot be seen as a final solution, we have been using git for two years now and we have recently published a fully reproducible article, which demonstrates the effectiveness of our proposal.

6.2. Performance Characterization and Optimization of IOs

In high-performance computing environments, parallel file systems provide a shared storage infrastructure to applications. In the situation where multiple applications access this shared infrastructure concurrently, their performance can be impaired because of interference. In [22], we improve improve performance by alleviating interference effects through a smart I/O scheduler scheduler that organizes and optimizes the applications' requests and adjusts the access pattern to the device characteristics. We apply machine learning techniques to automatically select the best scheduling algorithm for each situation. Our approach improves performance by up to 75

In [33], we present a new storage device profiling tool that characterizes the sequential to random throughput ratio for reads and writes of different sizes. As we explained previously, several optimizations aim at adapting applications' access patterns in order to generate contiguous accesses for improved performance when accessing storage devices like hard disks. However, when considering other storage options like RAID arrays and SSDs, the access time ratio between contiguous and non-contiguous accesses may not compensate for these optimizations' cost. In this scenario, the information provided by our tool could be used to dynamically decide if optimizations are beneficial for performance, which is why we took a particular attention to obtain accurate information in a minimal benchmarking time.

6.3. Application of Game Theory and Distributed Optimization to Wireless Networks

In wireless networks, channel conditions and user quality of service (QoS) requirements vary, often quite arbitrarily, with time (e.g. due to user mobility, fading, etc.) and users only have a very limited information about their environment. In such context optimizing transmission while taking power consumption into account is extremely challenging. We apply game theory technique to MIMO wireless network using OFDM or OFDMA where multi-path channels can be handled efficiently

In [25], [9], we show that distributed power allocation in heterogeneous OFDMA cognitive radio networks can be modeled as a game where each user equipment in the network engages in a non-cooperative game and allocates its available transmit power over subcarriers to maximize its individual utility. The corresponding equilibrium (Debreu, an extension of Nash Equilibrium) can be characterized with fractional programming and we provide sufficient conditions for computing such equilibria as fixed points of a water-filling best response operator. Using such approach can however be quite slow and is very sensitive to delay and information uncertainty (it may not converge). Therefore, we explain in [17] how signal covariance matrices in Gaussian MIMO multiple access channel can be learnt in presence of imperfect (and possibly delayed) feedback. The algorithm we propose is based on on the method of matrix exponential learning (MXL) and it has the same information and computation requirements as distributed water-filling. However our algorithm converge much faster even for large numbers of users and/or antennas per user and in the presence of user update asynchrony, random delays and/or ergodically changing channel conditions. Yet, since the system may evolve over time in an unpredictable fashion (e.g. due to changes in the wireless medium or the users' QoS requirements), static solution concepts (such as Nash equilibrium) may be no longer relevant and users must adapt to changes in the environment "on the fly", without being able to predict the system's evolution ahead of time. Hence, we focus on the concept of no-regret : policies that perform at least as well as the best fixed transmit profile in hindsight. In [31] and [41], we provide a formulation of power control as an online optimization problem and we show that the FM dynamics lead to no regret in this dynamic context. In [40] we apply this approach energy efficient

transmission in MIMO-OFDM systems and we show through numerical simulations that, in realistic network environments even under rapidly changing channel conditions, users can track their individually optimum transmit profile, achieving gains of up to 600in energy efficiency over uniform power allocation policies.

We also apply this technique to multi-carrier cognitive radio systems. Such systems allow opportunistic secondary users (SUs) to access portions of the spectrum that are unused by the network's licensed primary users (PUs), provided that the induced interference does not compromise the PUs' performance guarantees. In [14], we introduce a flexible spectrum access pricing schemes such that the corresponding Nash equilibrium is unique under very mild assumptions and satisfies the performance constraints. In addition, we derive a dynamic power allocation policy that converges to equilibrium within a few iterations (even for large numbers of users) and that relies only on local—and possibly imperfect—signal-to-interference-and-noise ratio measurements. In [24], we draw on exponential learning techniques to design an algorithm that is able to adapt to system changes "on the fly", i.e. such that the proposed transmit policy leads to no regret even under rapidly changing network conditions.

6.4. General Results in Game Theory

Our work on game theory is often motivated by applications to wireless networks but can often have a more general application.

In [38], motivated by applications to multi-antenna wireless networks, we propose a distributed and asynchronous algorithm for stochastic semidefinite programming This algorithm is a stochastic approximation of a continous-time matrix exponential scheme regularized by the addition of an entropy-like term to the problem's objective function. We show that the resulting algorithm converges almost surely to an (ϵ)-approximation of the optimal solution requiring only an unbiased estimate of the gradient of the problem's stochastic objective.

As explained in the previous section, classical Nash equilibrium concepts become irrelevant in situations where the environment evolves over time. In [15], we study one of the main concept of online learning and sequential decision problem known as regret minimization. Our objective is to provide a quick overview and a comprehensive introduction to online learning and game theory.

In practice, it is rarely reasonable to assume that players have access to the strategy of the others and implementing a best response can thus become cumbersome. Replicator dynamics is a fundamental approach in evolutionary game theory in which players adjust their strategies based on their actions' cumulative payoffs over time – specifically, by playing mixed strategies that maximize their expected cumulative payoff.

- In [19], we investigate the impact of payoff shocks on the evolution of large populations of myopic players that employ simple strategy revision protocols such as the "imitation of success". In the noiseless case, this process is governed by the standard (deterministic) replicator dynamics; in the presence of noise however, the induced stochastic dynamics are different from previous versions of the stochastic replicator dynamics (such as the aggregate-shocks model of Fudenberg and Harris, 1992). In this context, we show that strict equilibria are always stochastically asymptotically stable, irrespective of the magnitude of the shocks; on the other hand, in the high-noise regime, non-equilibrium states may also become stochastically asymptotically stable and dominated strategies may survive in perpetuity (they become extinct if the noise is low). Such behavior is eliminated if players are less myopic and revise their strategies based on their cumulative payoffs. In this case, we obtain a second order stochastic dynamical system whose attracting states coincide with the game's strict equilibria and where dominated strategies become extinct (a.s.), no matter the noise level.
- In [13], we study a new class of continuous-time learning dynamics consisting of a replicator-like drift adjusted by a penalty term that renders the boundary of the game's strategy space repelling. These penalty-regulated dynamics are equivalent to players keeping an exponentially discounted aggregate of their ongoing payoffs and then using a smooth best response to pick an action based on these performance scores. Building on the duality with evolutionary game theory, we design a discrete-time, payoff-based learning algorithm that converges to (arbitrarily precise) approximations of Nash equilibria in potential games. Moreover, the algorithm remains robust in the presence of

stochastic perturbations and observation errors, and it does not require any synchronization between players, which is a very important property when applying such technique to traffic engineering.

- In [18], we investigate an other class of reinforcement learning dynamics in which the players strategy adjustment is regularized with a strongly convex penalty term. In contrast to the class of penalty functions used to define smooth best responses in models of stochastic fictitious play, the regularizers used in this paper need not be infinitely steep at the boundary of the simplex. Dropping this requirement gives rise to an important dichotomy between steep and non-steep cases. In this general setting, our main results extend several properties of the replicator dynamics such as the elimination of dominated strategies, the asymptotic stability of strict Nash equilibria and the convergence of time-averaged trajectories to interior Nash equilibria in zero-sum games.
- In [37], we study a general class of game-theoretic learning dynamics in the presence of random payoff disturbances and observation noise, and we provide a unified framework that extends several rationality properties of the (stochastic) replicator dynamics and other game dynamics. In the unilateral case, we show that the stochastic dynamics under study lead to no regret, irrespective of the noise level. In the multi-player case, we find that dominated strategies become extinct (a.s.) and strict Nash equilibria remain stochastically asymptotically stable – again, independently of the perturbations' magnitude. Finally, we establish an averaging principle for 2-player games and we show that the empirical distribution of play converges to Nash equilibrium in zero-sum games under any noise level.

6.5. Simulation

Simgrid is a toolkit providing core functionalities for the simulation of distributed applications in heterogeneous distributed environments. Although it was initially designed to study large distributed computing environments such as grids, we have recently applied it to performance prediction of HPC configurations.

- Indeed, multi-core architectures comprising several GPUs have become mainstream but obtaining the maximum performance of such heterogeneous machines is challenging as it requires to carefully offload computations and manage data movements between the different processing units. The most promising and successful approaches so far build on task-based runtimes that abstract the machine and rely on opportunistic scheduling algorithms. As a consequence, the problem gets shifted to choosing the task granularity, task graph structure, and optimizing the scheduling strategies. Trying different combinations of these different alternatives is also itself a challenge. Indeed, getting accurate measurements requires reserving the target system for the whole duration of experiments. Furthermore, observations are limited to the few available systems at hand and may be difficult to generalize. In [21], we show how we crafted a coarse-grain hybrid simulation/emulation of StarPU, a dynamic runtime for hybrid architectures, over SimGrid. This approach allows to obtain performance predictions of classical dense linear algebra kernels accurate within a few percents and in a matter of seconds, which allows both runtime and application designers to quickly decide which optimization to enable or whether it is worth investing in higher-end GPUs or not. Additionally, it allows to conduct robust and extensive scheduling studies in a controlled environment whose characteristics are very close to real platforms while having reproducible behavior. In [30], we have extended this approach to the simulation of a multithreaded multifrontal QR solver of sparse matrices: QR-MUMPS. In our approach, the target high-end machines are calibrated only once to derive sound performance models. These models can then be used at will to quickly predict and study in a reproducible way the performance of such irregular and resource-demanding applications using solely a commodity laptop. Our approach also allows to study the memory consumption along time, which is a critical factor for such applications.
- Beside the inherent heterogeneity of distributed computing infrastructures, storage is also a essential component to cope with the tremendous increase in scientific data production and the ever-growing need for data analysis and preservation. Understanding the performance of a storage subsystem or dimensioning it properly is an important concern for which simulation can help. In [29], we detail

how we have extended SimGrid with storage simulation capacities and we list several concrete use cases of storage simulations in clusters, grids, clouds, and data centers for which the proposed extension would be beneficial.

$\Psi^2$is a simulation software of markovian models that is able to provide a perfect sampling of the stationary distribution. In [12], we consider open Jackson networks with losses with mixed finite and infinite queues and analyze the efficiency of sampling from their exact stationary distribution. We show that perfect sampling is possible, although the underlying Markov chain may have an infinite state space. The main idea is to use a Jackson network with infinite buffers (that has a product form stationary distribution) to bound the number of initial conditions to be considered in the coupling from the past scheme. We also provide bounds on the sampling time of this new perfect sampling algorithm for acyclic or hyper-stable networks. These bounds show that the new algorithm is considerably more efficient than existing perfect samplers even in the case where all queues are finite. We illustrate this efficiency through numerical experiments. We also extend our approach to variable service times and non-monotone networks such as queueing networks with negative customers.

6.6. Asymptotic Models

Analyzing a set of n stochastic entities interacting with each others can be particularly difficult but the *mean field approximation* is a very effective technique to characterize the probability distribution of such systems when the number of entities n grows very large. The limit system is generally deterministic and characterized by a differential equation that is more amenable to analysis and optimization. Such approximation however typically requires that the dynamics of the entities depend only on their state (the state space of each object does not scale with n the number of objects) but neither on their identity nor on their spatial location.

- In [28], we analyze a family of list-based cache replacement algorithms. We present explicit expressions for the cache content distribution and miss probability under some assumptions and we develop an algorithm with a time complexity that is polynomial in the cache size and linear in the number of items to compute the exact miss probability. We further introduce a mean field model to approximate the transient behavior of the miss probability and prove that this model becomes exact as the cache size and number of items tends to infinity. We show that the set of ODEs associated to the mean field model has a unique fixed point that can be used to approximate the miss probability in case the exact computation becomes too time consuming. Using this approximation, we provide guidelines on how to select a replacement algorithm within the family considered such that a good trade-off is achieved between the cache reactivity and its steady-state hit probability
- For distributed systems where /locality/ is essential in the dynamics the mean-field approach requires to resort to discretization of space into a finite number of cells to fit in the classical framework. Such approach not only scales badly but also requires that spatial interactions are weak. One of the tool to tackle this difficult problem comes from statistical physics and is popular in biology: pair approximation. In [26], we successfully apply this approach to the "Power of Two Choice" load balancing paradigm: each incoming task is allocated to the least loaded of two servers picked at random among a collection of n servers. We study the power of two-choice in a setting where the two servers are not picked independently at random but are connected by an edge in an underlying graph. Our problem is motivated by systems in which choices are geometrically constrained (e.g., a bike-sharing system). We study a dynamic setting in which jobs leave the system after being served by a server to which is was allocated. Our focus is when each server has few neighbors (typically 2 to 4) for which an mean-field approximation is not accurate. We build the pair-approximation equations and show that they describe accurately the steady-state of the system. Our results show that, even in a graph of degree 2, choosing between two neighboring improve dramatically the performance compared to a random allocation.
- In [8], we consider a queueing system composed of a dispatcher that routes deterministically jobs to a set of non-observable queues working in parallel. In this setting, the fundamental problem is which policy should the dispatcher implement to minimize the stationary mean waiting time of the incoming jobs. We present a structural property that holds in the classic scaling of the system where

the network demand (arrival rate of jobs) grows proportionally with the number of queues. Assuming that each queue of type r is replicated k times, we consider a set of policies that are periodic with period $k \sum_r p_r$ and such that exactly p_r jobs are sent in a period to each queue of type r. When $k \to \infty$, our main result shows that all the policies in this set are equivalent, in the sense that they yield the same mean stationary waiting time, and optimal, in the sense that no other policy having the same aggregate arrival rate to all queues of a given type can do better in minimizing the stationary mean waiting time. Furthermore, the limiting mean waiting time achieved by our policies is a convex function of the arrival rate in each queue, which facilitates the development of a further optimization aimed at solving the fundamental problem above for large systems.

6.7. Trace and Statistical Analysis

Although we often use Markovian approaches to model large scale distributed system, these probabilistic tools can also be used to lay the foundation of statistical analysis of traces of real systems.

- In [36], we explain how we apply statistical statistical modelling and statistical inference of the ANR GEOMEDIA corpus, that is a collection of international RSS news feeds. Central to this project, RSS news feeds are viewed as a representation of the information in geopolitical space. As such they allow us to study media events of global extent and how they affect international relations. Here we propose hidden Markov models (HMM) as an adequate modelling framework to study the evolution of media events in time. This set of models respect the characteristic properties of the data, such as temporal dependencies and correlations between feeds. Its specific structure corresponds well to our conceptualisation of media attention and media events. We specify the general model structure that we use for modelling an ensemble of RSS news feeds. Finally, we apply the proposed models to a case study dedicated to the analysis of the media attention for the Ebola epidemic which spread through West Africa in 2014.
- The use of stochastic formalisms, such as Stochastic Automata Networks (SAN), can be very useful for statistical prediction and behavior analysis. Once well fitted, such formalisms can generate probabilities about a target reality. These probabilities can be seen as a statistical approach of knowledge discovery. However, the building process of models for real world problems is time consuming even for experienced modelers. Furthermore, it is often necessary to be a domain specialist to create a model. In [34], we present a new method to automatically learn simple SAN models directly from a data source. This method is encapsulated in a tool called SAN GEnerator (SANGE). Through examples we show how this new model fitting method is powerful and relatively easy to use, which can grant access to a much broader community to such powerful modeling formalisms.
- In [32], we have presented our recent results on macroscopic analysis of huge traces of parallel/distributed applications. To identify a *macroscopic phenomenon* over large traces, one needs to change the representation scale and to aggregate data both in time, space and application structure through meaningful operators to propose *multi-scale visualizations*. The question is then to know the quantity of information lost by such scaling to be able to correctly interpret them. The principles underlying this approach are based on information theory since the conditional entropy of an aggregation indicates the quantity of information loss when data are aggregated. This approach has been integrated in the Framesoc framework [35].
- In [27], We study the problem of making forecasts about the future availability of bicycles in stations of a bike-sharing system (BSS). This is relevant in order to make recommendations guaranteeing that the probability that a user will be able to make a journey is sufficiently high. To this end, we use probabilistic predictions obtained from a queuing theoretical time-inhomogeneous model of a BSS. The model is parametrized and successfully validated using historical data from the Vélib ' BSS of the City of Paris. We develop a critique of the standard root-mean-square-error (RMSE), commonly adopted in the bike-sharing research as an index of the prediction accuracy, because it does not account for the stochasticity inherent in the real system. Instead we introduce a new metric

based on scoring rules. We evaluate the average score of our model against classical predictors used in the literature. We show that these are outperformed by our model for prediction horizons of up to a few hours. We also discuss that, in general, measuring the current number of available bikes is only relevant for prediction horizons of up to few hours.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry: Alcatel Lucent-Bell

A common laboratory between Inria and the Alcatel Lucent-Bell Labs was created in early 2008 and consists on three research groups (ADR). MESCAL leads the ADR on self-optimizing networks (SELFNET). The researchers involved in this project are Bruno Gaujal and Panayotis Mertikopoulos.

7.2. Bilateral Contracts with Industry: Stimergy

Stimergy is a startup that aims at developing a distributed data center built by connecting mini data centers embedded in digital boilers installed in multi-unit residential buildings. Each boiler contains several servers and the dissipated power can thus be used to cover a large part of the annual energy requirements for preparing domestic hot water for a building. Such infrastructure drastically reduces the energy required to operate data centers, while reducing total cost of infrastructure and ownership. Mescal (Olivier Richard, and Michael Mercier, full-time Inria engineer) provides the necessary expertise for the realization and implementation of software infrastructure allowing the coordination of operating such mini data center.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. CIMENT

The CIMENT project (Intensive Computing, Numerical Modeling and Technical Experiments, http://ciment.ujf-grenoble.fr/) gathers a wide scientific community involved in numerical modeling and computing (from numerical physics and chemistry to astrophysics, mechanics, bio-modeling and imaging) and the distributed computer science teams from Grenoble. Several heterogeneous distributed computing platforms were set up (from PC clusters to IBM SP or alpha workstations) each being originally dedicated to a scientific domain. More than 600 processors are available for scientific computation. The MESCAL project-team provides expert skills in high performance computing infrastructures. The members of MESCAL involved in this project are Pierre Neyron and Olivier Richard.

8.1.2. Cluster Région

Partners: the Inria GRAAL project-team, the LSR-IMAG and IN2P3-LAPP laboratories.

The MESCAL project-team is a member of the regional "cluster" project on computer science and applied mathematics, the focus of its participation is on handling large amount of data large scale architecture.

8.2. National Initiatives

8.2.1. Inria Large Scale Initiative

- *HEMERA, 2010-2014* Leading action "Completing challenging experiments on Grid'5000 (Methodology)" (see https://www.grid5000.fr/Hemera).

  Experimental platforms like Grid'5000 or PlanetLab provide an invaluable help to the scientific community, by making it possible to run very large-scale experiments in controlled environment. However, while performing relatively simple experiments is generally easy, it has been shown that the complexity of completing more challenging experiments (involving a large number of nodes, changes to the environment to introduce heterogeneity or faults, or instrumentation of the platform to extract data during the experiment) is often underestimated.

  This working group explores different complementary approaches, that are the basic building blocks for building the next level of experimentation on large scale experimental platforms.

8.2.2. ANR

- *ANR GAGA (2014-2017)*

  GAGA is a "Young Researchers" project funded by the French National Research Agency (ANR) to explore the Geometric Aspects of GAmes. The GAGA teamis spread over three different locations in France (Paris, Toulouse and Grenoble), and is coordinated by Vianney Perchet, assistant professor (Maître de Conférences) in the Probabilities and Random Models laboratory in Université Paris VII.

  As the name suggests, our project's focus is game theory, a rapidly developing subject with growing applications in economics, social sciences, computer science, engineering, evolutionary biology, etc. As it turns out, many game theoretical topics and tools have a strong geometrical or topological flavor: the structure of a game's equilibrium set, the design of equilibrium-computing algorithms, Blackwell approachability, the geometric character of the replicator dynamics, the use of semi-algebraicity concepts in stochastic games, and many others. Accordingly, our objective is to perform a systematic study of these geometric aspects of game theory and, by so doing, to establish new links between areas that so far appeared unrelated (such as Hessian-Riemannian geometry and discrete choice theory).

- *ANR MARMOTE, 2013-2016*. Partners: Inria Sophia (MAESTRO), Inria Rocquencourt (DIOGEN), PRiSM laboratory from University of Versailles-Saint-Quentin, Telecom SudParis (SAMOVAR), University Paris-Est Créteil (*Spécification et vérification de systèmes*), Université Pierre-et-Marie-Curie/LIP6.

  The project aims at realizing a software prototype dedicated to Markov chain modeling. It gathers seven teams that will develop advanced resolution algorithms and apply them to various domains (reliability, distributed systems, biology, physics, economy).

- *ANR NETLEARN, 2013-2015*. Partners: PRiSM laboratory from University of Versailles-Saint-Quentin, Telecom ParisTech, Orange Labs, LAMSADE/University Paris Dauphine, Alcatel-Lucent, Inria (MESCAL).

  The main objective of the project is to propose a novel approach of distributed, scalable, dynamic and energy efficient algorithms for managing resources in a mobile network. This new approach relies on the design of an orchestration mechanism of a portfolio of algorithms. The ultimate goal of the proposed mechanism is to enhance the user experience, while at the same time to better utilize the operator resources. User mobility and new services are key elements to take into account if the operator wants to improve the user quality of experience. Future autonomous network management and control algorithms will thus have to deal with a real-time dynamicity due to user mobility and to traffic variations resulting from various usages. To achieve this goal, we focus on two central aspects of mobile networks (the management of radio resources at the Radio Access Network level and the management of the popular contents users want to get access to) and intend to design distributed

learning mechanisms in non-stationary environments, as well as an orchestration mechanism that applies the best algorithms depending on the situation.

- *ANR SONGS, 2012-2015*. Partners: Inria Nancy (Algorille), Inria Sophia (MASCOTTE), Inria Bordeaux (CEPAGE, HiePACS, RunTime), Inria Lyon (AVALON), University of Strasbourg, University of Nantes.

  The last decade has brought tremendous changes to the characteristics of large scale distributed computing platforms. Large grids processing terabytes of information a day and the peer-to-peer technology have become common even though understanding how to efficiently exploit such platforms still raises many challenges. As demonstrated by the USS SimGrid project funded by the ANR in 2008, simulation has proved to be a very effective approach for studying such platforms. Although even more challenging, we think the issues raised by petaflop/exaflop computers and emerging cloud infrastructures can be addressed using similar simulation methodology.

  The goal of the SONGS project (Simulation of Next Generation Systems) is to extend the applicability of the SimGrid simulation framework from grids and peer-to-peer systems to clouds and high performance computation systems. Each type of large-scale computing system will be addressed through a set of use cases and led by researchers recognized as experts in this area.

  Any sound study of such systems through simulations relies on the following pillars of simulation methodology: Efficient simulation kernel; Sound and validated models; Simulation analysis tools; Campaign simulation management.

8.2.3. National Organizations

Jean-Marc Vincent is member of the scientific committees of the CIST (Centre International des Sciences du Territoire).

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

8.3.1.1. Mont-Blanc 2

Program: FP7 Programme

Project acronym: Mont-Blanc 2

Project title: Mont-Blanc: European scalable and power efficient HPC platform based on low-power embedded technology

Duration: October 2013 - September 2016

Coordinator: BSC (Barcelone)

Other partners: BULL - Bull SAS (France), STMicroelectronics - (GNB SAS) (France), ARM - (United Kingdom), JUELICH - (Germany), BADW-LRZ - (Germany), USTUTT - (Germany), CINECA - (Italy), CNRS - (France), Inria - (France), CEA - (France), UNIVERSITY OF BRISTOL - (United Kingdom), ALLINEA SW LIM - (United Kingdom)

Abstract: Energy efficiency is already a primary concern for the design of any computer system and it is unanimously recognized that future Exascale systems will be strongly constrained by their power consumption. This is why the Mont-Blanc project has set itself the following objective: to design a new type of computer architecture capable of setting future global High Performance Computing (HPC) standards that will deliver Exascale performance while using 15 to 30 times less energy. Mont-Blanc 2 contributes to the development of extreme scale energy-efficient platforms, with potential for Exascale computing, addressing the challenges of massive parallelism, heterogeneous computing, and resiliency. Mont-Blanc 2 has great potential to create new market opportunities for successful EU technology, by placing embedded architectures in servers and HPC.

The Mont-Blanc 2 proposal has 4 objectives:

1. To complement the effort on the Mont-Blanc system software stack, with emphasis on programmer tools (debugger, performance analysis), system resiliency (from applications to architecture support), and ARM 64-bit support.
2. To produce a first definition of the Mont-Blanc Exascale architecture, exploring different alternatives for the compute node (from low-power mobile sockets to special-purpose high-end ARM chips), and its implications on the rest of the system.
3. To track the evolution of ARM-based systems, deploying small cluster systems to test new processors that were not available for the original Mont-Blanc prototype (both mobile processors and ARM server chips).
4. To provide continued support for the Mont-Blanc consortium, namely operations of the Mont-Blanc prototype, and hands-on support for our application developers

8.3.1.2. QUANTICOL

Program: The project is a member of Fundamentals of Collective Adaptive Systems (FOCAS), a FET-Proactive Initiative funded by the European Commission under FP7.

Project acronym: QUANTICOL

Project title: A Quantitative Approach to Management and Design of Collective and Adaptive Behaviours

Duration: 04 2013 – 03 2017

Coordinator: Jane Hillston (University of Edinburgh, Scotland)

Other partners: University of Edinburgh (Scotland); Istituto di Scienza e Tecnologie della Informazione (Italy); IMT Lucca (Italy) and University of Southampton (England).

Abstract: The main objective of the QUANTICOL project is the development of an innovative formal design framework that provides a specification language for collective adaptive systems (CAS) and a large variety of tool-supported, scalable analysis and verification techniques. These techniques will be based on the original combination of recent breakthroughs in stochastic process algebras and associated verification techniques, and mean field/continuous approximation and control theory. Such a design framework will provide scalable extensive support for the verification of developed models, and also enable and facilitate experimentation and discovery of new design patterns for emergent behaviour and control over spatially distributed CAS.

8.3.1.3. NEWCOM#

Program: FP7-ICT-318306

Project acronym: NEWCOM#

Project title: Network of Excellence in Wireless Communications

Duration: 11 2012 – 10 2015

Coordinator: Consorzio Nazionale Interuniversitario per le Telecomunicazioni (Italy)

Other partners: Aalborg Universitet (AAU). Denmark; Bilkent Üniversitesi (Bilkent). Turkey; Centre National de la Recherche Scientifique (CNRS). France; Centre Tecnològic de Telecomunicacions de Catalunya (CTTC). Spain; Institute of Accelerating Systems and Applications (IASA). Greece; Inesc Inovacao; Instituto de Novas Tecnologias (INOV). Portugal; Poznan University of Technology (PUT). Poland; Technion - Israel Institute of Technology (TECHNION). Israel; Technische Universitaet Dresden (TUD). Germany; University of Cambridge (UCAM). United Kingdom; Universite Catholique de Louvain (UCL). Belgium; Oulun Yliopisto (UOULU). Finland

Abstract: NEWCOM# is a project funded under the umbrella of the 7th Framework Program of the European Commission (FP7-ICT-318306). NEWCOM# pursues long-term, interdisciplinary research on the most advanced aspects of wireless communications like Finding the Ultimate Limits of Communication Networks, Opportunistic and Cooperative Communications, or Energy- and Bandwidth-Efficient Communications and Networking.

8.3.1.4. HPC4E

Title: HPC for Energy

Program: H2020

Duration: 01 2016 – 01 2018

Coordinator: Barcelona Supercomputing Center

Inria contact: Stephane Lanteri

Other partners:

- Europe: Lancaster University (ULANC), Centro de Investigaciones Energéticas Medioambientales y Tecnológicas (CIEMAT), Repsol S.A. (REPSOL), Iberdrola Renovables Energía S.A. (IBR), Total S.A. (TOTAL).
- Brazil: Fundação Coordenação de Projetos, Pesquisas e Estudos Tecnoclógicos (COPPE), National Laboratory for Scientific Computation (LNCC), Instituto Tecnológico de Aeronáutica (ITA), Petroleo Brasileiro S. A. (PETROBRAS), Universidade Federal do Rio Grande do Sul (INF-UFRGS), Universidade Federal de Pernambuco (CER-UFPE)

Abstract: The main objective of the HPC4E project is to develop beyond-the-state-of-the-art high performance simulation tools that can help the energy industry to respond future energy demands and also to carbon-related environmental issues using the state-of-the-art HPC systems. The other objective is to improve the cooperation between energy industries from EU and Brazil and the cooperation between the leading research centres in EU and Brazil in HPC applied to energy industry. The project includes relevant energy industrial partners from Brazil and EU, which will benefit from the project's results. They guarantee that TRL of the project technologies will be very high. This includes sharing supercomputing infrastructures between Brazil and EU. The cross-fertilization between energy-related problems and other scientific fields will be beneficial at both sides of the Atlantic.

8.3.2. Collaborations in European Programs, except FP7 & H2020

8.3.2.1. CROWN

Program: European Community and Greek General Secretariat for Research and Technology

Project acronym: CROWN

Project title: Optimal Control of Self Organized Wireless Networks

Duration: 2012-2015

Coordinator: Tassiulas Leandros

Other partners: Thales, University of Thessaly, National and Kapodistrian University of Athens, Athens University of Economics and Business

Abstract: Wireless networks are rapidly becoming highly complex systems with large numbers of heterogeneous devices interacting with each other, often in a harsh environment. In the absence of central control, network entities need to self-organize to reach an efficient operating state, while operating in a distributed fashion. Depending on whether the operating criteria are individual or global, nodes interact in an autonomic or coordinated way. Despite recent progress in autonomic networks, the fundamental understanding of the operational behaviour of large-scale networks is still lacking. This project will address these emergent network properties, by introducing new tools and concepts from other disciplines.

We will first analyze how imperfect network state information can be harvested and distributed efficiently through the network using machine learning techniques. We will design flexible methodologies to shape the competition between autonomous nodes for resources, with aim to maintain robust social optimality. Both cooperating and non-cooperating game-theoretic models will be used. We also consider networks with nodes coordinating to achieve a joint task, e.g., global optimization. Using algorithms inspired from statistical physics, we will address two representative paradigms in

the context of wireless ad hoc networks, namely connectivity optimization and the localization of a network of primary sources from a sensor network.

Finally, we will explore delay tolerant networks as a case study of an emerging class of networks that, while sharing most of the characteristics of traditional autonomic or coordinated networks, they present unique challenges, due to the intermittency and constant fluctuations of the connectivity. We will study tradeoffs involving delay, the impact of mobility on information transfer, and the optimal usage of resources by using tools from information theory and stochastic evolution theory.

8.3.3. Collaborations with Major European Organizations

University of Athens: Panayotis Mertikopoulos was an invited professor for 3 months.

EPFL: Laboratoire pour les communications informatiques et leurs applications 2, Institut de systèmes de communication ISC, Ecole polytechnique fédérale de Lausanne (Switzerland). We collaborate with Jean-Yves Leboudec (EPFL) and Pierre Pinson (DTU) on electricity markets.

University of Edinburgh and Istituto di Scienza e Tecnologie della Informazione: we strongly collaborate through the Quanticol European project.

University of Antwerp: we collaborate with Benny Van Houdt on caching problems.

TU Wien: Research Group Parallel Computing, Technische Universität Wien (Austria). We collaborate with Sascha Hunold on experimental methodology and reproducibility of experiments in HPC.

8.4. International Initiatives

8.4.1. Inria International Labs

8.4.1.1. North America

- JLESC (former JLPC) (Joint Laboratory for Extreme-Scale Computing) with University of University of Illinois Urbana Champaign, Argonne Nat. Lab and BSC. Several members of MESCAL are partners of this laboratory, and have done several visits to Urbana-Champaign or NCSA.

8.4.2. Inria Associate Teams not involved in an Inria International Labs

8.4.2.1. EXASE

Title: Exascale Computing Scheduling and Energy

International Partner (Institution - Laboratory - Researcher):

Universidade Federal do Rio Grande do Sul (Brazil) - INF (INF) - Nicolas MAILLARD

Start year: 2014

See also: https://team.inria.fr/exase/

The main scientific goal of this collaboration for the three years is the development of state-of-the-art energy-aware scheduling algorithms for exascale systems. Three complementary research directions have been identified : (1) Fundamentals for the scaling of schedulers: develop new scheduling algorithms for extreme exascale machines and use existing workloads to validate the proposed scheduling algorithms (2) Design of schedulers for large-scale infrastructures : propose energy-aware schedulers in large-scale infrastructures and develop adaptive scheduling algorithms for exascale machines (3) Tools for the analysis of large scale schedulers : develop aggregation methodologies for scheduler analysis to propose synthetized visualizations for large traces analysis and then analyze schedulers and energy traces for correlation analysis

8.4.3. Inria International Partners

8.4.3.1. Declared Inria International Partners

- MESCAL has strong connections with both UFRGS (Porto Alegre, Brazil) and USP (Sao Paulo, Brazil). The creation of the LICIA common laboratory (see next section) has made this collaboration even tighter.

- MESCAL has strong bounds with the University of Illinois Urbana Champaign, within the (Joint Laboratory on Petascale Computing, see previous section).

8.4.4. Participation In other International Programs

8.4.4.1. South America

- LICIA. The CNRS, Inria, the Universities of Grenoble, Grenoble INP and Universidade Federal do Rio Grande do Sul have created the LICIA (*Laboratoire International de Calcul intensif et d'Informatique Ambiante*). Jean-Marc Vincent is the director of the laboratory, on the French side.

  The main themes are high performance computing, language processing, information representation, interfaces and visualization as well as distributed systems.

  More information can be found at http://www.inf.ufrgs.br/licia/.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

Stan Zachary and James Cruise, from Heriot-Watt University at Edinburgh, came for a week in the context of the European Quanticol project. Lucas Schnorr and Philippe Navaux from UFRGS (Porto Alegre, Brazil) both came for a week in the context of the EXASE associated team.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

Bruno Gaujal organized the Game Theory In'Tech seminar on June 2015.

9.1.2. Scientific events selection

9.1.2.1. Member of the conference program committees

Arnaud Legrand was in the PC of HiPC and co-organized the RepPar workshop (Workshop on Reproducibility in Parallel Computing).

The members of the team regularly review numerous papers for international conferences.

9.1.3. Journal

9.1.3.1. Reviewer - Reviewing activities

The members of the MESCAL team regularly review articles for JPDC, DAM, IEEE Transactions on Networking/Automatic Control/Cloud Computing/Parallel and Distributed Computing/..., FGCS, ParCo, ...

9.1.4. Invited talks

Bruno Gaujal was invited to the Bacceli seminar in Paris [1]. Bruno Gaujal was invited to give a keynote at the Congrès of the "Société des Mathématiques Appliquées et Industrielles". Bruno Gaujal and Jean-Marc Vincent were invited to participate and present their work to the Symposium honoring Erol Gelenbe.

[1] http://www.di.ens.fr/~blaszczy/FB60/

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Master/PhD: Panayotis Mertikopoulos, "Game Theory for the Working Economist", 35 Eq. TD, University of Athens, Department of Economic Sciences PhD program (UADPhilEcon), Athens, Greece.

Master : Bruno Gaujal, Discrete Event Systems, 18 h, (M2R), MPRI, Paris.

Master : Bruno Gaujal, Advanced Performance evaluation, 9 h, (M2), Ensimag, Grenoble.

Master: Arnaud Legrand, Parallel Systems, 21 h, (M2R), Mosig.

Master: Arnaud Legrand and Jean-Marc Vincent, Performance Evaluation, 15 h, (M2R), Mosig.

Master: Arnaud Legrand and Jean-Marc Vincent, Probability and simulation, performance evaluation 72 h, (M1), RICM, Polytech Grenoble.

Master: Jean-Marc Vincent, Mathematics for computer science, 18 h , (M1) Mosig.

Master: Jean-Marc Vincent, workshop on the methodology in computer science research, 15 h, (M2R) Mosig.

Master : Olivier Richard, Networking, 40 Eq. TD, (M1), RICM, Polytech Grenoble

Master : Olivier Richard, Physical Computing Eq. 60 TD, L1 and M1, Joseph Fourier University and Polytech Grenoble

DU: Jean-Marc Vincent, Informatique et sciences du numérique, 20 h, (Professeurs de lycée).

E-learning

SPOC: Jean-Marc Vincent, Informatique et sciences du numérique, 6 mois, plate-forme pairformnce,rectorat de Grenoble, Professeurs de lycée, formation continue, environ 50 inscrits.

9.2.2. Supervision

PhD : Luka Stanisic, *A Reproducible Research Methodology for Designing and Conducting Faithful Simulations of Dynamic Task-based Scientific Applications* Defended Oct. 30, 2014[6], supervised by Arnaud Legrand and Jean-François Méhaut.

HdR : Arnaud Legrand, *Scheduling for Large Scale Distributed Computing Systems: Approaches and Performance Evaluation Issues*, Université Grenoble Alpes, Defended Nov. 2[5], 2015.

9.2.3. Juries

- Bruno Gaujal was a reviewer for the PhD of Josu Doncel, on “Efficiency of Distributed Queueing Games and of Path Discovery Algorithms“ at University of Toulouse, on March 30th 2015
- Bruno Gaujal was a reviewer for the HDR of Nidhi Hegde.

9.3. Popularization

- fête de la science: atelier sciences manuelles du numérique.
- MathC2+: sciences manuelles du numérique.
- Interventions dans des classes de seconde sur la simulation simcity.
- Participation à l’IREM de Grenoble, groupe algorithmique.

10. Bibliography

Major publications by the team in recent years

[1] H. CASANOVA, A. GIERSCH, A. LEGRAND, M. QUINSON, F. SUTER. *Versatile, Scalable, and Accurate Simulation of Distributed Applications and Platforms*, in "Journal of Parallel and Distributed Computing", June 2014, vol. 74, n^{o} 10, pp. 2899-2917 [*DOI :* 10.1016/J.JPDC.2014.06.008], https://hal.inria.fr/hal-01017319

[2] N. GAST, B. GAUJAL, J.-Y. LE BOUDEC. *Mean field for Markov decision processes: from discrete to continuous optimization*, in "Automatic Control, IEEE Transactions on", 2012, vol. 57, n° 9, pp. 2266–2280

[3] B. JAVADI, D. KONDO, J.-M. VINCENT, D. P. ANDERSON. *Discovering Statistical Models of Availability in Large Distributed Systems: An Empirical Study of SETI@home*, in "IEEE Transactions on Parallel and Distributed Systems", 2010

[4] P. MERTIKOPOULOS, E. V. BELMEGA, A. L. MOUSTAKAS, S. LASAULCE. *Distributed Learning Policies for Power Allocation in Multiple Access Channels*, in "IEEE JSAC", January 2012, vol. 30, n° 1, pp. 96-106

Publications of the year

Doctoral Dissertations and Habilitation Theses

[5] A. LEGRAND. *Scheduling for Large Scale Distributed Computing Systems: Approaches and Performance Evaluation Issues*, Université Grenoble Alpes, November 2015, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-01247932

[6] L. STANISIC. *A Reproducible Research Methodology for Designing and Conducting Faithful Simulations of Dynamic Task-based Scientific Applications*, Université Grenoble Alpes, October 2015, https://tel.archives-ouvertes.fr/tel-01248109

Articles in International Peer-Reviewed Journals

[7] A. ANCEL, I. ASSENMACHER, K. I. BABA, J. CISONNI, Y. FUJISO, P. GONÇALVES, M. IMBERT, K. KOYAMADA, P. NEYRON, N. KAZUNORI, O. HIROYUKI, A.-C. ORGERIE, X. PELORSON, B. RAFFIN, N. SAKAMOTO, E. SAKANE, S. WADA, S. SHIMOJO, A. VAN HIRTUM. *PetaFlow: a global computing-networking-visualisation unitwith social impact*, in "International Research Journal of Computer Science", April 2015, vol. 2, n° 4, https://hal.inria.fr/hal-01231826

[8] J. ANSELMI, B. GAUJAL, T. NESTI. *Control of parallel non-observable queues: asymptotic equivalence and optimality of periodic policies*, in "stochastic systems", December 2015, vol. 5, n° 1 [*DOI :* 10.1214/14-SSY146], https://hal.archives-ouvertes.fr/hal-01102936

[9] G. BACCI, E. V. BELMEGA, P. MERTIKOPOULOS, L. SANGUINETTI. *Energy-Aware Competitive Power Allocation for Heterogeneous Networks Under QoS Constraints*, in "IEEE Transactions on Wireless Communications", 2015, vol. 14, n° 9, pp. 4728-4742, https://hal.inria.fr/hal-01073494

[10] D. BORGETTO, R. CHAKODE, B. DEPARDON, C. EICHLER, J.-M. GARCIA, H. HBAIEB, T. MONTEIL, E. PELORCE, A. RACHDI, A. AL SHEIKH, P. STOLF. *Hybrid approach for energy aware management of multi-cloud architecture integrating user machines*, in "Journal of Grid Computing", 2015, https://hal.archives-ouvertes.fr/hal-01228290

[11] T. BUCHERT, C. RUIZ, L. NUSSBAUM, O. RICHARD. *A survey of general-purpose experiment management tools for distributed systems*, in "Future Generation Computer Systems", 2015, vol. 45, pp. 1 - 12 [*DOI :* 10.1016/J.FUTURE.2014.10.007], https://hal.inria.fr/hal-01087519

[12] A. BUSIC, S. DURAND, B. GAUJAL, F. PERRONNIN. *Perfect sampling of Jackson queueing networks*, in "Queueing Systems", 2015, vol. 80, n° 3, 37 p. [*DOI :* 10.1007/S11134-015-9436-Z], https://hal.inria.fr/hal-01236542

[13] P. COUCHENEY, B. GAUJAL, P. MERTIKOPOULOS. *Penalty-Regulated Dynamics and Robust Learning Procedures in Games*, in "Mathematics of Operations Research", 2015, vol. 40, n° 3, pp. 611-633 [*DOI :* 10.1287/MOOR.2014.0687], https://hal.inria.fr/hal-01235243

[14] S. D'ORO, P. MERTIKOPOULOS, A. L. MOUSTAKAS, S. PALAZZO. *Interference-based pricing for opportunistic multi-carrier cognitive radio systems*, in "IEEE Transactions on Wireless Communications", 2015, vol. 14, n° 12, pp. 6536 - 6549 [*DOI :* 10.1109/TWC.2015.2456063], https://hal.archives-ouvertes.fr/hal-01239593

[15] M. FAURE, P. GAILLARD, B. GAUJAL, V. PERCHET. *Online Learning and Game Theory. A quick overview with recent results and applications* , in "ESAIM: Proceedings", October 2015 [*DOI :* 10.1051/PROC/201551014], https://hal.inria.fr/hal-01237039

[16] R. LARAKI, P. MERTIKOPOULOS. *Inertial game dynamics and applications to constrained optimization*, in "SIAM Journal on Control and Optimization", 2015, vol. 53, n° 5, pp. 3141-3170, 31 pages, 4 figures, https://hal.inria.fr/hal-00920928

[17] P. MERTIKOPOULOS, A. L. MOUSTAKAS. *Learning in an uncertain world: MIMO covariance matrix optimization with imperfect feedback*, in "IEEE Transactions on Signal Processing", 2016, vol. 64, n° 1, pp. 5-18 [*DOI :* 10.1109/TSP.2015.2477053], https://hal.archives-ouvertes.fr/hal-01239585

[18] P. MERTIKOPOULOS, W. H. SANDHOLM. *Learning in games via reinforcement and regularization*, in "Mathematics of Operations Research", 2015, https://hal.archives-ouvertes.fr/hal-01239590

[19] P. MERTIKOPOULOS, Y. VIOSSAT. *Imitation dynamics with payoff shocks*, in "International Journal of Game Theory", 2015, https://hal.inria.fr/hal-01099014

[20] L. STANISIC, A. LEGRAND, V. DANJEAN. *An Effective Git And Org-Mode Based Workflow For Reproducible Research*, in "Operating Systems Review", 2015, vol. 49, pp. 61 - 70 [*DOI :* 10.1145/2723872.2723881], https://hal.inria.fr/hal-01112795

[21] L. STANISIC, S. THIBAULT, A. LEGRAND, B. VIDEAU, J.-F. MÉHAUT. *Faithful Performance Prediction of a Dynamic Task-Based Runtime System for Heterogeneous Multi-Core Architectures*, in "Concurrency and Computation: Practice and Experience", May 2015, 16 p. [*DOI :* 10.1002/CPE], https://hal.inria.fr/hal-01147997

[22] F. ZANON BOITO, R. V. KASSICK, P. O. A. NAVAUX, Y. DENNEULIN. *Automatic I/O scheduling algorithm selection for parallel file systems*, in "Concurrency and Computation: Practice and Experience", 2015 [*DOI :* 10.1002/CPE.3606], https://hal.inria.fr/hal-01247942

International Conferences with Proceedings

[23] J. ASSUNÇÃO, P. FERNANDES, L. LOPES, S. NORMEY. *A dimensionality reduction process to forecast events through stochastic models*, in "International Conference on Software Engineering and Knowledge Engineering", Pittsburgh, United States, July 2015, https://hal.inria.fr/hal-01247905

[24] E. V. BELMEGA, P. MERTIKOPOULOS. *Energy-efficient power allocation in dynamic multi-carrier systems*, in "VTC Spring 2015: Proceedings of the 2015 IEEE Vehicular Technology Conference", Glasgow, United Kingdom, 2015, https://hal.archives-ouvertes.fr/hal-01239592

[25] S. D'ORO, P. MERTIKOPOULOS, A. L. MOUSTAKAS, S. PALAZZO. *Cost-efficient power allocation in OFDMA cognitive radio networks*, in "EUCNC' 15: Proceedings of the 2015 European Conference on Networks and Communications", Paris, France, 2015, https://hal.archives-ouvertes.fr/hal-01239586

[26] N. GAST. *The Power of Two Choices on Graphs: the Pair-Approximation is Accurate*, in "Workshop on MAthematical performance Modeling and Analysis", Portland, United States, June 2015, https://hal.inria.fr/hal-01199271

[27] N. GAST, G. MASSONNET, D. REIJSBERGEN, M. TRIBASTONE. *Probabilistic Forecasts of Bike-Sharing Systems for Journey Planning*, in "The 24th ACM International Conference on Information and Knowledge Management (CIKM 2015)", Melbourne, Australia, October 2015 [*DOI :* 10.1145/2806416.2806569], https://hal.inria.fr/hal-01185840

[28] N. GAST, B. VAN HOUDT. *Transient and Steady-state Regime of a Family of List-based Cache Replacement Algorithms*, in "ACM SIGMETRICS 2015", Portland, United States, June 2015 [*DOI :* 10.1145/2745844.2745850], https://hal.inria.fr/hal-01143838

[29] A. LÈBRE, A. LEGRAND, F. SUTER, P. VEYRE. *Adding Storage Simulation Capacities to the SimGrid Toolkit: Concepts, Models, and API*, in "CCGrid 2015 - Proceedings of the 15th IEEE/ACM Symposium on Cluster, Cloud and Grid Computing", Shenzhen, China, IEEE/ACM, May 2015, pp. 251-260 [*DOI :* 10.1109/CCGRID.2015.134], https://hal.inria.fr/hal-01197128

[30] L. STANISIC, E. AGULLO, A. BUTTARI, A. GUERMOUCHE, A. LEGRAND, F. LOPEZ, B. VIDEAU. *Fast and Accurate Simulation of Multithreaded Sparse Linear Algebra Solvers*, in "The 21st IEEE International Conference on Parallel and Distributed Systems", Melbourne, Australia, The 21st IEEE International Conference on Parallel and Distributed Systems, December 2015, https://hal.inria.fr/hal-01180272

[31] I. STIAKOGIANNAKIS, P. MERTIKOPOULOS, C. TOUATI. *No more tears: A no-regret approach to power control in dynamically varying MIMO networks*, in "WiOpt '15 - Proceedings of the 13th International Symposium and Workshops on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks", Mumbai, India, May 2015, pp. 467 - 474 [*DOI :* 10.1109/WIOPT.2015.7151107], https://hal.archives-ouvertes.fr/hal-01239588

Conferences without Proceedings

[32] J.-M. VINCENT. *Macroscopic analysis of huge traces of parallel/distributed applications*, in "ISCIS 30th International Symposium on Computer and Information Sciences. Symposium Honouring Professor Erol Gelenbe", London, United Kingdom, September 2015, https://hal.inria.fr/hal-01247911

[33] F. ZANON BOITO, R. KASSICK, P. O. A. NAVAUX, Y. DENNEULIN. *Towards fast profiling of storage devices regarding access sequentiality*, in "30th Annual ACM Symposium on Applied Computing", Salamanca, Spain, April 2015, pp. 2015–2020 [*DOI :* 10.1145/2695664.2695701], https://hal.inria.fr/hal-01247938

Research Reports

[34] J. ASSUNÇÃO, P. FERNANDES, L. LOPES, A. STUDENY, J.-M. VINCENT. *SANGE -Stochastic Automata Networks Generator. A tool to efficiently predict events through structured Markovian models (extended version)*, Inria Rhône-Alpes ; Grenoble University ; Pontifícia Universidade Católica do Rio Grande do Sul ; Inria, March 2015, n^{o} RR-8724, https://hal.inria.fr/hal-01149604

[35] G. PAGANO, V. MARANGOZOVA-MARTIN. *Effective Data Management for Interactive Trace Analysis*, Inria - Research Centre Grenoble – Rhône-Alpes ; Inria, March 2015, n^{o} RT-0460, 26 p. , https://hal.inria.fr/hal-01155518

[36] A. STUDENY, R. LAMARCHE-PERRIN, J.-M. VINCENT. *Studying Media Events through Spatio-Temporal Statistical Analysis*, Inria Grenoble - Rhone-Alpes, September 2015, https://hal.inria.fr/hal-01246239

Other Publications

[37] M. BRAVO, P. MERTIKOPOULOS. *On the Robustness of Learning in Games with Stochastically Perturbed Payoff Observations*, January 2015, working paper or preprint, https://hal.inria.fr/hal-01098494

[38] B. GAUJAL, P. MERTIKOPOULOS. *A stochastic approximation algorithm for stochastic semidefinite programming*, December 2015, 25 pages, 4 figures, https://hal.archives-ouvertes.fr/hal-01239587

[39] K. GEORGIEV, V. MARANGOZOVA-MARTIN. *Facing the Challenge of Nondeterminism in MPSoC Debugging*, January 2015, Submitted to the JSA Elsevier Journal, https://hal.inria.fr/hal-01103620

[40] P. MERTIKOPOULOS, E. V. BELMEGA. *Learning to be green: robust energy efficiency maximization in dynamic MIMO-OFDM systems*, December 2015, 25 pages, 4 figures, https://hal.archives-ouvertes.fr/hal-01239591

[41] I. STIAKOGIANNAKIS, P. MERTIKOPOULOS, C. TOUATI. *Adaptive Power Allocation and Control in Time-Varying Multi-Carrier MIMO Networks*, December 2015, 25 pages, 4 figures, https://hal.archives-ouvertes.fr/hal-01239589