IN PARTNERSHIP WITH:
**CNRS**

**Ecole normale supérieure de Lyon**

**Université Claude Bernard (Lyon 1)**

# Activity Report 2016

# Project-Team AVALON

## Algorithms and Software Architectures for Distributed and HPC Platforms

IN COLLABORATION WITH: Laboratoire de l'Informatique du Parallélisme (LIP)

RESEARCH CENTER
**Grenoble - Rhône-Alpes**

THEME
**Distributed and High Performance Computing**

# Table of contents

# Project-Team AVALON

*Creation of the Team: 2012 February 01, updated into Project-Team: 2014 July 01*

**Keywords:**

**Computer Science and Digital Science:**

1. - Architectures, systems and networks
1.1.1. - Multicore
1.1.4. - High performance computing
1.1.5. - Exascale
1.1.6. - Cloud
1.1.7. - Peer to peer
1.1.13. - Virtualization
1.2.1. - Dynamic reconfiguration
1.3. - Distributed Systems
1.6. - Green Computing
2.1.6. - Concurrent programming
2.1.7. - Distributed programming
2.1.10. - Domain-specific languages
2.5.2. - Component-based Design
2.6.2. - Middleware
3.1.2. - Data management, quering and storage
3.1.3. - Distributed data
3.1.8. - Big data (production, storage, transfer)
3.1.9. - Database
3.2.1. - Knowledge bases
4. - Security and privacy
4.4. - Security of equipment and software
4.5. - Formal methods for security
6.2.7. - High performance computing
7.1. - Parallel and distributed algorithms
7.3. - Optimization
7.11. - Performance evaluation

**Other Research Topics and Application Domains:**

1.1.9. - Bioinformatics
3.2. - Climate and meteorology
3.4. - Risks
3.4.2. - Industrial risks and waste
4.1. - Fossile energy production (oil, gas)
4.2.2. - Fusion
4.5. - Energy consumption
4.5.1. - Green computing
6.1.1. - Software engineering

8.1.1. - Energy for smart buildings
9.4.1. - Computer science
9.6. - Reproducibility

# 1. Members

**Research Scientists**

Christian Perez [Team leader, Inria, Senior Researcher, HDR]
Marcos Dias de Assunção [Inria, Starting Research position]
Gilles Fedak [Inria, Researcher, HDR]
Thierry Gautier [Inria, Researcher]
Laurent Lefevre [Inria, Researcher, HDR]
Frédéric Suter [CNRS, Researcher, HDR]

**Faculty Members**

Eddy Caron [ENS Lyon, Associate Professor, HDR]
Yves Caniou [Univ. Lyon I, Associate Professor]
Jean-Patrick Gelas [Univ. Lyon I, Associate Professor]
Olivier Glück [Univ. Lyon I, Associate Professor]

**Engineers**

Jean-Christophe Mignot [CNRS, Engineer Researcher]
Simon Delamare [CNRS, Engineer Researcher]
Matthieu Imbert [Inria, Engineer (40%)]
Mathilde Boutigny [Inria, since Oct 2016]
Romaric Guillier [CNRS, until Jun 2016]
Salem Harrache [Inria, until Feb 2016]
Arnaud Lefray [Inria]
Olivier Mornard [Inria, until Jul 2016]
Marc Pinhede [Inria, until Apr 2016]
Mathieu Veyrand [Inria, since Nov 2016]

**PhD Students**

Daniel Balouek-Thomert [New Generation SR]
Radu Carpa [ENS Lyon]
Hadrien Croubois [ENS Lyon]
Alexandre Da Silva Veith [Univ. Lyon, since Oct 2016]
Jad Darrous [Inria, since Oct 2016]
Pedro de Souza Bento Da Silva [Inria]
Vincent Lanore [ENS Lyon, until Aug 2016]
Valentin Lorentz [ENS Lyon]
Issam Rais [Inria]
Jérôme Richard [Inria]
Violaine Villebonnet [Inria]

**Post-Doctoral Fellow**

Hélène Coullon [Inria, until Sep 2016]

**Administrative Assistant**

Evelyne Blesle [Inria]

**Others**

David Loup [INSA Lyon / Inria part-time apprentice]
Thomas Pellissier-Tanon [ENS Lyon, Long scholarship since Sep 2016]

# 2. Overall Objectives

## 2.1. Presentation

The fast evolution of hardware capabilities in terms of wide area communication, computation and machine virtualization leads to the requirement of another step in the abstraction of resources with respect to parallel and distributed applications. Those large scale platforms based on the aggregation of large clusters (Grids), huge datacenters (Clouds), collections of volunteer PC (Desktop computing platforms), or high performance machines (Supercomputers) are now available to researchers of different fields of science as well as to private companies. This variety of platforms and the way they are accessed also have an important impact on how applications are designed (*i.e.,* the programming model used) as well as how applications are executed (*i.e.,* the runtime/middleware system used). The access to these platforms is driven through the use of different services providing mandatory features such as security, resource discovery, virtualization, load-balancing, monitoring, etc.

The goal of the Avalon team is to execute parallel and/or distributed applications on parallel and/or distributed resources while ensuring user and system objectives with respect to performance, cost, energy, security, etc. Users are not interested in the resources used during the execution. Instead, they are interested in how their application is going to be executed: in which duration, at which cost, what is the environmental footprint involved, etc. This vision of utility computing has been strengthened by the cloud concepts and by the short lifespan of supercomputers (around three years) compared to application lifespan (tens of years). Therefore, a major issue is to design models, systems, and algorithms to execute applications on resources while ensuring user constraints (price, performance, etc.) as well as system administrator constraints (maximizing resource usage, minimizing energy consumption, etc.).

## 2.2. Objectives

To achieve the vision proposed in Section 2.1, the Avalon project aims at making progress to four complementary research axes: energy, data, component models, and application scheduling.

### *2.2.1. Energy Application Profiling and Modelization*

Avalon will improve the profiling and modeling of scientific applications with respect to energy consumption. In particular, it will require to improve the tools that measure the energy consumption of applications, virtualized or not, at large scale, so as to build energy consumption models of applications.

### *2.2.2. Data-intensive Application Profiling, Modeling, and Management*

Avalon will improve the profiling, modeling, and management of scientific applications with respect to CPU and data intensive applications. The challenges are to improve the performance prediction of parallel regular applications, to model and simulate (complex) intermediate storage components, and data-intensive applications, and last to deal with data management for hybrid computing infrastructures.

### *2.2.3. Resource-Agnostic Application Description Model*

Avalon will design component-based models to capture the different facets of parallel and distributed applications while being resource agnostic, so that they can be optimized for a particular execution. In particular, the proposed component models will integrate energy and data modeling results. Avalon in particular targets OpenMP runtime as a specific use case.

### *2.2.4. Application Mapping and Scheduling*

Avalon will propose multi-criteria mapping and scheduling algorithms to meet the challenge of automatizing the efficient utilization of resources taking into consideration criteria such as performance (CPU, network, and storage), energy consumption, and security. Avalon will in particular focus on application deployment, workflow applications, and security management in clouds.

All our theoretical results will be validated with software prototypes using applications from different fields of science such as bioinformatics, physics, cosmology, etc. The experimental testbed GRID'5000 (cf Section 5.9) will be our platform of choice for experiments.

# 3. Research Program

## 3.1. Energy Application Profiling and Modelization

International roadmaps schedule to build exascale systems by the 2018 time frame. According to the Top500 list published in November 2013, the most powerful supercomputer is the Tianhe-2 platform, a machine with more than 3,000,000 cores. It consumes more than 17 MW for a maximum performance of 33 PFlops while the Defense Advanced Research Projects Agency (DARPA) has set to 20 MW the maximum energy consumption of an exascale supercomputer [40].

Energy efficiency is therefore a major challenge for building next generation large scale platforms. The targeted platforms will gather hundreds of million cores, low power servers, or CPUs. Besides being very important, their power consumption will be dynamic and irregular.

Thus, to consume energy efficiently, we aim at investigating two research directions. First, we need to improve the measure, the understanding, and the analysis of the large-scale platform energy consumption. Unlike approaches [41] that mix the usage of internal and external wattmeters on a small set of resources, we target high frequency and precise internal and external energy measurements of each physical and virtual resources on large scale distributed systems.

Secondly, we need to find new mechanisms that consume less and better on such platforms. Combined with hardware optimizations, several works based on shutdown or slowdown approaches aim at reducing energy consumption of distributed platforms and applications. To consume less, we first plan to explore the provision of accurate estimation of the energy consumed by applications without pre-executing and knowing them while most of the works try to do it based on in-depth application knowledge (code instrumentation [44], phase detection for specific HPC applications [49], etc.). As a second step, we aim at designing a framework model that allows interactions, dialogues and decisions taken in cooperation between the user/application, the administrator, the resource manager, and the energy supplier. While smart grid is one of the last killer scenarios for networks, electrical provisioning of next generation large IT infrastructures remains a challenge.

## 3.2. Data-intensive Application Profiling, Modeling, and Management

Recently, the term “Big Data” has emerged to design data sets or collections so large that they become intractable for classical tools. This term is most of the time implicitly linked to “analytics” to refer to issues such as curation, storage, search, sharing, analysis, and visualization. However, the Big Data challenge is not limited to data-analytics, a field that is well covered by programming languages and run-time systems such as Map-Reduce. It also encompasses data-intensive applications. These applications can be sorted into two categories. In High Performance Computing (HPC), data-intensive applications leverage post-petascale infrastructures to perform highly parallel computations on large amount of data, while in High Throughput Computing (HTC), a large amount of independent and sequential computations are performed on huge data collections.

These two types of data-intensive applications (HTC and HPC) raise challenges related to profiling and modeling that the Avalon team proposes to address. While the characteristics of data-intensive applications are very different, our work will remain coherent and focused. Indeed, a common goal will be to acquire a better understanding of both the applications and the underlying infrastructures running them to propose the best match between application requirements and infrastructure capacities. To achieve this objective, we will extensively rely on logging and profiling in order to design sound, accurate, and validated models. Then, the proposed models will be integrated and consolidated within a single simulation framework (SIMGRID). This will allow us to explore various potential “what-if?” scenarios and offer objective indicators to select interesting infrastructure configurations that match application specificities.

Another challenge is the ability to mix several heterogeneous infrastructures that scientists have at their disposal (*e.g.,* Grids, Clouds, and Desktop Grids) to execute data-intensive applications. Leveraging the aforementioned results, we will design strategies for efficient data management service for hybrid computing infrastructures.

## 3.3. Resourc-Agnostic Application Description Model

When programming in parallel, users expect to obtain performance improvement, whatever the cost is. For long, parallel machines have been simple enough to let a user program them given a minimal abstraction of their hardware. For example, MPI [43] exposes the number of nodes but hides the complexity of network topology behind a set of collective operations; OpenMP [47] simplifies the management of threads on top of a shared memory machine while OpenACC [46] aims at simplifying the use of GPGPU.

However, machines and applications are getting more and more complex so that the cost of manually handling an application is becoming very high [42]. Hardware complexity also stems from the unclear path towards next generations of hardware coming from the frequency wall: multi-core CPU, many-core CPU, GPGPUs, deep memory hierarchy, etc. have a strong impact on parallel algorithms. Hence, even though an abstract enough parallel language (UPC, Fortress, X10, etc.) succeeds, it will still face the challenge of supporting distinct codes corresponding to different algorithms corresponding to distinct hardware capacities.

Therefore, the challenge we aim to address is to define a model, for describing the structure of parallel and distributed applications that enables code variations but also efficient executions on parallel and distributed infrastructures. Indeed, this issue appears for HPC applications but also for cloud oriented applications. The challenge is to adapt an application to user constraints such as performance, energy, security, etc.

Our approach is to consider component based models [50] as they offer the ability to manipulate the software architecture of an application. To achieve our goal, we consider a “compilation” approach that transforms a resource-agnostic application description into a resource-specific description. The challenge is thus to determine a component based model that enables to efficiently compute application mapping while being tractable. In particular, it has to provide an efficient support with respect to application and resource elasticity, energy consumption and data management. OpenMP runtime is a specific use case that we target.

## 3.4. Application Mapping and Scheduling

This research axis is at the crossroad of the Avalon team. In particular, it gathers results of the three other research axis. We plan to consider application mapping and scheduling through the following three issues.

### *3.4.1. Application Mapping and Software Deployment*

Application mapping and software deployment consist in the process of assigning distributed pieces of software to a set of resources. Resources can be selected according to different criteria such as performance, cost, energy consumption, security management, etc. A first issue is to select resources at application launch time. With the wide adoption of elastic platforms, *i.e.,* platforms that let the number of resources allocated to an application to be increased or decreased during its execution, the issue is also to handle resource selection at runtime.

The challenge in this context corresponds to the mapping of applications onto distributed resources. It will consist in designing algorithms that in particular take into consideration application profiling, modeling, and description.

A particular facet of this challenge is to propose scheduling algorithms for dynamic and elastic platforms. As the amount of elements can vary, some kind of control of the platforms must be used accordingly to the scheduling.

### 3.4.2. Non-Deterministic Workflow Scheduling

Many scientific applications are described through workflow structures. Due to the increasing level of parallelism offered by modern computing infrastructures, workflow applications now have to be composed not only of sequential programs, but also of parallel ones. New applications are now built upon workflows with conditionals and loops (also called non-deterministic workflows).

These workflows can not be scheduled beforehand. Moreover cloud platforms bring on-demand resource provisioning and pay-as-you-go billing models. Therefore, there is a problem of resource allocation for non-deterministic workflows under budget constraints and using such an elastic management of resources.

Another important issue is data management. We need to schedule the data movements and replications while taking job scheduling into account. If possible, data management and job scheduling should be done at the same time in a closely coupled interaction.

### 3.4.3. Security Management in Cloud Infrastructure

Security has been proven to be sometimes difficult to obtain [48] and several issues have been raised in Clouds. Nowadays virtualization is used as the sole mechanism to secure different users sharing resources on Clouds. But, due to improper virtualization of all the components of Clouds (such as micro-architectural components), data leak and modification can occur. Accordingly, next-generation protection mechanisms are required to enforce security on Clouds and provide a way to cope with the current limitation of virtualization mechanisms.

As we are dealing with parallel and distributed applications, security mechanisms must be able to cope with multiple machines. Our approach is to combine a set of existing and novel security mechanisms that are spread in the different layers and components of Clouds in order to provide an in-depth and end-to-end security on Clouds. To do it, our first challenge is to define a generic model to express security policies.

Our second challenge is to work on security-aware resource allocation algorithms. The goal of such algorithms is to find a good trade-off between security and unshared resources. Consequently, they can limit resources sharing to increase security. It leads to complex trade-off between infrastructure consolidation, performance, and security.

# 4. Application Domains

## 4.1. Overview

The Avalon team targets applications with large computing and/or data storage needs, which are still difficult to program, maintain, and deploy. Those applications can be parallel and/or distributed applications, such as large scale simulation applications or code coupling applications. Applications can also be workflow-based as commonly found in distributed systems such as grids or clouds.

The team aims at not being restricted to a particular application field, thus avoiding any spotlight. The team targets different HPC and distributed application fields, which bring use cases with different issues. This will be eased by our various collaborations: the team participates to the INRIA-Illinois Joint Laboratory for Petascale Computing, the Physics, Radiobiology, Medical Imaging, and Simulation French laboratory of excellence, the E-Biothon project, the INRIA large scale initiative Computer and Computational Sciences at Exascale (C2S@Exa), and to BioSyL, a federative research structure about Systems Biology of the University of Lyon. Moreover, the team members have a long tradition of cooperation with application developers such as CERFACS and EDF R&D. Last but not least, the team has a privileged connection with CC IN2P3 that opens up collaborations, in particular in the astrophysics field.

In the following, some examples of representative applications we are targeting are presented. In addition to highlighting some application needs, they also constitute some of the use cases we will use to valide our theoretical results.

## 4.2. Climatology

The world's climate is currently changing due to the increase of the greenhouse gases in the atmosphere. Climate fluctuations are forecasted for the years to come. For a proper study of the incoming changes, numerical simulations are needed, using general circulation models of a climate system. Simulations can be of different types: HPC applications (*e.g.,* the NEMO framework [45] for ocean modelization), code-coupling applications (*e.g.,* the OASIS coupler [51] for global climate modeling), or workflows (long term global climate modeling).

As for most applications the team is targeting, the challenge is to thoroughly analyze climate-forecasting applications to model their needs in terms of programing model, execution model, energy consumption, data access pattern, and computing needs. Once a proper model of an application has been set up, appropriate scheduling heuristics could be designed, tested, and compared. The team has a long tradition of working with CERFACS on this topic, for example in the LEGO (2006-09) and SPADES (2009-12) French ANR projects.

## 4.3. Astrophysics

Astrophysics is a major field to produce large volume of data. For instance, the Large Synoptic Survey Telescope (http://www.lsst.org/lsst/) will produce 15 TB of data every night, with the goals of discovering thousands of exoplanets and of uncovering the nature of dark matter and dark energy in the universe. The Square Kilometer Array (http://www.skatelescope.org/) produces 9 Tbits/s of raw data. One of the scientific projects related to this instrument called Evolutionary Map of the Universe is working on more than 100 TB of images. The Euclid Imaging Consortium will generate 1 PB data per year.

Avalon collaborates with the *Institut de Physique Nucléaire de Lyon* (IPNL) laboratory on large scale numerical simulations in astronomy and astrophysics. Contributions of the Avalon members have been related to algorithmic skeletons to demonstrate large scale connectivity, the development of procedures for the generation of realistic mock catalogs, and the development of a web interface to launch large cosmological simulations on GRID'5000.

This collaboration, that continues around the topics addressed by the CLUES project (http://www.clues-project.org), has been extended thanks to the tight links with the CC-IN2P3. Major astrophysics projects execute part of their computing, and store part of their data on the resources provided by the CC-IN2P3. Among them, we can mention SNFactory, Euclid, or LSST. These applications constitute typical use cases for the research developed in the Avalon team: they are generally structured as workflows and a huge amount of data (from TB to PB) is involved.

## 4.4. Bioinformatics

Large-scale data management is certainly one of the most important applications of distributed systems in the future. Bioinformatics is a field producing such kinds of applications. For example, DNA sequencing applications make use of MapReduce skeletons.

The Avalon team is a member of BioSyL (http://www.biosyl.org), a Federative Research Structure attached to University of Lyon. It gathers about 50 local research teams working on systems biology. Moreover, the team cooperated with the French Institute of Biology and Chemistry of Proteins (IBCP http://www.ibcp.fr) in particular through the ANR MapReduce project where the team focuses on a bio-chemistry application dealing with protein structure analysis. Avalon have also starting working with the Inria Beagle team (https://team.inria.fr/beagle/) on artificial evolution and computational biology as the challenges are around high performance computation and data management.

# 5. New Software and Platforms

## 5.1. Active Data

**Participants:** Gilles Fedak [correspondant], Valentin Lorentz, Laurent Lefevre.

FUNCTIONAL DESCRIPTION

Active Data is a free software system that tracks the life cycle of data distributed across heterogeneous software and infrastructures.

As the volume of data grows exponentially, the management of these data becomes more complex in proportion. A key point is to handle the complexity of the Data Life Cycle, *i.e.,* the various operations performed on data: transfer, archiving, replication, deletion, etc. Indeed, data-intensive applications span over a large variety of devices and e-infrastructures which implies that many systems are involved in data management and processing. Active Data is a new approach to automate and improve the expressiveness of data management applications. Active Data consists of a formal model that captures the essential data life cycle stages and properties: creation, deletion, replication, derivation, transient unavailability, uniform naming, and many more. Active Data provides a programming model that simplify the development of data life cycle management applications. Active Data allows code execution at each stage of the data life cycle: routines provided by programmers are executed when a set of events (creation, replication, transfer, deletion) happen to any data.

- URL: http://active-data.gforge.inria.fr

## 5.2. DIET

**Participants:** Daniel Balouek-Thomert, Yves Caniou, Eddy Caron, Arnaud Lefray.

FUNCTIONAL DESCRIPTION

DIET (Distributed Interactive Engineering Toolbox) is a middleware designed for high-performance computing in a heterogeneous and distributed environment (workstations, clusters, grids, clouds). Three main developments are done in 2016:

Proxmox support: DIET has a new Cloud extension that can be used to interact with this PVE (Proxmox Virtual Environment) solution. Proxmox is a complete open source server virtualization management software, based on KVM virtualization and container-based virtualization. It manages KVM virtual machines, Linux containers (LXC), storage, virtualized networks, and HA clusters. This extension is a contribution provided by the NewGeneration-SR company.

DaaS support: The goal of this development was to add Data-as-a-Service (DaaS) functionality to the DIET middleware via an optional module. We have added the GDConnect plug-in to the DIET Data manager that provides the capability to deal with Google's Cloud Storage.

Energy Support: We designed, implementation and evaluation of an energy-efficient resource management system that builds upon DIET and NSDE-divisible tasks with precedence constraints.

- Partners: CNRS - ENS Lyon - UCBL Lyon 1
- Contact: Eddy Caron
- URL: http://graal.ens-lyon.fr/diet/

## 5.3. Execo

**Participant:** Matthieu Imbert.

FUNCTIONAL DESCRIPTION

Execo offers a Python API for asynchronous control of local or remote, standalone or parallel, unix processes. It is especially well suited for quickly and easily scripting workflows of parallel/distributed operations on local or remote hosts: automate a scientific workflow, conduct computer science experiments, perform automated tests, etc. The core python package is execo. The execo_g5k package provides a set of tools and extensions for the Grid'5000 testbed. The execo_engine package provides tools to ease the development of computer sciences experiments. Execo is used directly by 15 to 30 users, in and out of the Avalon team, and is also used by a few Grid'5000 tools such as kwapi, funk, topo5k, g5k_bench_flops.

- Contact: Matthieu Imbert
- URL: http://execo.gforge.inria.fr

## 5.4. Kaapi

**Participant:** Thierry Gautier.

Kaapi is a library for high performance applications running on multi-cores/multi- processors with support for multi-GPUs. Kaapi provides ABI compliant implementations of libGOMP (GCC runtime for OpenMP) and libomp.so (CLANG and Intel compiler). It was one of the target runtime of the K'Star compiler (http://kstar.gforge.inria.fr).

Web site:

- Partners: EPI Corse (Philippe Virouleau, François Broquedis)
- Contact: Thierry Gautier
- URL: http://kaapi.gforge.inria.fr

## 5.5. Kwapi

**Participants:** Jean-Patrick Gelas, Laurent Lefevre.

FUNCTIONAL DESCRIPTION

Kwapi is a software framework dealing with energy monitoring of large scale infrastructures through heterogeneous energy sensors. Kwapi has been designed inside the FSN XLCloud project for OpenStack infrastructures. Through the support of Hemera Inria project, kwapi has been extended and deployed in production mode to support easy and large scale energy profiling of the Grid'5000 resources.

- Contact: Laurent Lefevre
- URL: https://launchpad.net/kwapi

## 5.6. L2C and DirectL2C

**Participants:** Hélène Coullon, Vincent Lanore, Christian Perez, Jérôme Richard.

FUNCTIONAL DESCRIPTION

L2C (http://hlcm.gforge.inria.fr) is a Low Level Component model implementation targeting at use-cases where overhead matters such as High-Performance Computing. L2C does not offer network transparency neither language transparency. Instead, L2C lets the user choose between various kinds of interactions between components, some with ultra low overhead and others that support network transport. L2C is extensible as additional interaction kinds can be added quite easily. L2C currently supports C++, FORTRAN 2013, MPI and CORBA interactions.

DirectL2C is an extension to L2C that enables efficient and consistent reconfiguration of large scale L2C based assemblies. It provides an assembly model enhanced with domains, tranformations, and transformation adapters.

- Partners: CEA/Maison de la Simulation (Julien Bigot)
- Contact: Christian Perez
- URL: http://hlcm.gforge.inria.fr/l2c:start

## 5.7. Sam4C

**Participants:** Eddy Caron, Arnaud Lefray, Marc Pinhede, Mathieu Veyrand.

SCIENTIFIC DESCRIPTION

This editor is generated in Java from an EMF -Eclipse Modeling Framework- metamodel to simplify any modifications or extensions. The application model and the associated security policy are compiled in a single XML file which serves as input for an external Cloud security-aware scheduler. Alongside with this editor, Cloud architecture models and provisioning algorithms are provided for simulation (in the current version) or real deployments (in future versions).

FUNCTIONAL DESCRIPTION

Sam4C (https://gforge.inria.fr/projects/sam4c/) -Security-Aware Models for Clouds- is a graphical and textual editor to model Cloud applications (as virtual machines, processes, files and communications) and describe its security policy. Sam4C is suitable to represent any static application without deadline or execution time such as n-tiers or parallel applications.

- Contact: Eddy Caron
- URL: https://gforge.inria.fr/projects/sam4c/

## 5.8. SimGrid

**Participant:** Frédéric Suter.

FUNCTIONAL DESCRIPTION

SimGrid is a toolkit that provides core functionalities for the simulation of distributed applications in heterogeneous distributed environments. The simulation engine uses algorithmic and implementation techniques toward the fast simulation of large systems on a single machine. The models are theoretically grounded and experimentally validated. The results are reproducible, enabling better scientific practices.

Its models of networks, cpus and disks are adapted to (Data)Grids, P2P, Clouds, Clusters and HPC, allowing multi-domain studies. It can be used either to simulate algorithms and prototypes of applications, or to emulate real MPI applications through the virtualization of their communication, or to formally assess algorithms and applications that can run in the framework.

The formal verification module explores all possible message interleavings in the application, searching for states violating the provided properties. We recently added the ability to assess liveness properties over arbitrary and legacy codes, thanks to a system-level introspection tool that provides a finely detailed view of the running application to the model checker. This can for example be leveraged to verify both safety or liveness properties, on arbitrary MPI code written in C/C++/Fortran.

- Partners: CNRS - ENS Rennes
- Contact: Martin Quinson (EPC Myriads)
- URL: http://simgrid.gforge.inria.fr/

## 5.9. Grid'5000

**Participants:** Laurent Lefevre, Simon Delamare, David Loup, Christian Perez.

FUNCTIONAL DESCRIPTION

The Grid'5000 experimental platform is a scientific instrument to support computer science research related to distributed systems, including parallel processing, high performance computing, cloud computing, operating systems, peer-to-peer systems and networks. It is distributed on 10 sites in France and Luxembourg, including Lyon. Grid'5000 is a unique platform as it offers to researchers many and varied hardware resources and a complete software stack to conduct complex experiments, ensure reproducibility and ease understanding of results. In 2016, a new cluster financially supported by Inria has been deployed on the Grid5000 Lyon site.

- Contact: Laurent Lefevre
- URL: https://www.grid5000.fr/

# 6. New Results

## 6.1. Energy Efficiency of Large Scale Distributed Systems

**Participants:** Laurent Lefevre, Daniel Balouek-Thomert, Eddy Caron, Radu Carpa, Marcos Dias de Assunção, Jean-Patrick Gelas, Olivier Glück, Jean-Christophe Mignot, Violaine Villebonnet.

### 6.1.1. Energy Efficient Core Networks with SDN

This work [14], [15] seeks to improve the energy efficiency of backbone networks by providing an intra-domain Software Defined Network (SDN) approach to selectively and dynamically turn off and on a subset of links. We proposed the STREETE framework (SegmenT Routing based Energy Efficient Traffic Engineering) that represents an online method to switch some links off/on dynamically according to the network load. We have implemented a working prototype in the OMNET++ simulator and design a validation platform [15] based on NetFPGA and Raspberry equipment with SDN frameworks (ONOS).

### 6.1.2. Energy Proportionality in HPC Systems

Energy savings are among the most important topics concerning Cloud and HPC infrastructures nowadays. Servers consume a large amount of energy, even when their computing power is not fully utilized. These static costs represent quite a concern, mostly because many datacenter managers are over-provisioning their infrastructures compared to the actual needs. This results in a high part of wasted power consumption. In this work [25], [24], [23] , we proposed the BML ("Big, Medium, Little") infrastructure, composed of heterogeneous architectures, and a scheduling framework dealing with energy proportionality. We introduce heterogeneous power processors inside datacenters as a way to reduce energy consumption when processing variable workloads. Our framework brings an intelligent utilization of the infrastructure by dynamically executing applications on the architecture that suits their needs, while minimizing energy consumption. Our first validation process focuses on distributed stateless web servers scenario and we analyze the energy savings achieved through energy proportionality. This research activity is performed with the collaboration of Sepia Team (IRIT, Toulouse) through the co-advising of Violaine Villebonnet.

### 6.1.3. Energy-Aware Server Provisioning

Several approaches to reduce the power consumption of datacenters have been described in the literature, most of which aim to improve energy efficiency by trading off performance for reducing power consumption. However, these approaches do not always provide means for administrators and users to specify how they want to explore such trade-offs. This work [11] provides techniques for assigning jobs to distributed resources, exploring energy efficient resource provisioning. We use middleware-level mechanisms to adapt resource allocation according to energy-related events and user-defined rules. A proposed framework enables developers, users and system administrators to specify and explore energy efficiency and performance trade-offs without detailed knowledge of the underlying hardware platform. Evaluation of the proposed solution under three scheduling policies shows gains of 25% in energy-efficiency with minimal impact on the overall application performance. We also evaluate reactivity in the adaptive resource provisioning This research activity is performed with the collaboration of NewGen SR society through the co-advising of Daniel Balouek-Thomert.

### 6.1.4. Improving Energy Re-Usage of Large Scale Computing Systems

The heat induced by computing resources is generally a waste of energy in supercomputers. This is especially true in very large scale supercomputers, where the produced heat has to be compensated with expensive and energy consuming cooling systems. Energy is a critical point for future supercomputing trends that currently try to achieve exascale, without having its energy consumption reaching an important fraction of a nuclear power plant. Thus, new ways of generating or recovering energy have to be explored. Energy harvesting consists in recovering wasted energy. ThermoElectric Generators (TEGs) aim to recover energy by converting wasted dissipated energy into usable electricity. By combining computing units (CU) and TEGs at very large scale, we spotted a potential way to recover energy from wasted heat generated by computations on supercomputers. In this work [30], [20], we study the potential gains in combining TEGs with computational units at petascale and exascale. We explored the technology behind TEGs, and finally our results concerning binding TEGs and computational units in a petascale and exascale system. With the available technology, we demonstrate that the use of TEGs in a supercomputer environment could be realistic and quickly profitable, and hence have a positive environmental impact.

## 6.2. MPI Application Simulation

**Participant:** Frédéric Suter.

### *6.2.1. The SMPI approach*

In [37], we summarized our recent work and developments on SMPI, a flexible simulator of MPI applications. In this tool, we took a particular care to ensure our simulator could be used to produce fast and accurate predictions in a wide variety of situations. Although we did build SMPI on SimGrid whose speed and accuracy had already been assessed in other contexts, moving such techniques to a HPC workload required significant additional effort. Obviously, an accurate modeling of communications and network topology was one of the key to such achievements. Another less obvious key was the choice to combine in a single tool the possibility to do both offline and online simulation.

## 6.3. MapReduce Computations on Hybrid Distributed Computations Infrastructures

**Participant:** Gilles Fedak.

### *6.3.1. Availability and Network-Aware MapReduce Task Scheduling over the Internet*

MapReduce offers an easy-to-use programming paradigm for processing large datasets. In our previous work, we have designed a MapReduce framework called BitDew-MapReduce for desktop grid and volunteer computing environment, that allows non-expert users to run data-intensive MapReduce jobs on top of volunteer resources over the Internet. However, network distance and resource availability have great impact on MapReduce applications running over the Internet. To address this, an availability and network-aware MapReduce framework over the Internet is proposed in [9]. Simulation results show that the MapReduce job response time could be decreased by 27.15%, thanks to Naive Bayes Classifier-based availability prediction and landmark-based network estimation.

## 6.4. Managing Big Data Life Cycle

**Participants:** Gilles Fedak, Valentin Lorentz, Laurent Lefevre.

### *6.4.1. Data Energy Traceability*

In this work, we have opened a new research topic around the energy traceability of data. The objective is to answer the question of how many energy has been consumed to produce a particular data. This work is partially based on the concept of data life cycle, that is extended to record each step of the data life cycle.

## 6.5. Desktop Grid Computing

**Participant:** Gilles Fedak.

### *6.5.1. Multi-Criteria and Satisfaction Oriented Scheduling for Hybrid Distributed Computing Infrastructures*

Assembling and simultaneously using different types of distributed computing infrastructures (DCI) like Grids and Clouds is an increasingly common situation. Because infrastructures are characterized by different attributes such as price, performance, trust, greenness, the task scheduling problem becomes more complex and challenging. In [7], we presented the design for a fault-tolerant and trust-aware scheduler, which allows to execute Bag-of-Tasks applications on elastic and hybrid DCI, following user-defined scheduling strategies. Our approach, named Promethee scheduler, combines a pull-based scheduler with multi-criteria Promethee decision making algorithm. Because multi-criteria scheduling leads to the multiplication of the possible scheduling strategies, we proposed SOFT, a methodology that allows to find the optimal scheduling strategies given a set of application requirements. The validation of this method is performed with a simulator that fully implements the Promethee scheduler and recreates an hybrid DCI environment including Internet Desktop

Grid, Cloud and Best Effort Grid based on real failure traces. A set of experiments shows that the Promethee scheduler is able to maximize user satisfaction expressed accordingly to three distinct criteria: price, expected completion time and trust, while maximizing the infrastructure useful employment from the resources owner point of view. Finally, we present an optimization which bounds the computation time of the Promethee algorithm, making realistic the possible integration of the scheduler to a wide range of resource management software.

## 6.6. HPC Component Models and OpenMP

**Participants:** Hélène Coullon, Vincent Lanore, Christian Perez, Jérôme Richard, Thierry Gautier.

### *6.6.1. Combining Both a Component Model and a Task-based Model*

We have studied the feasibility of efficiently combining both a software component model and a task-based model. Task based models are known to enable efficient executions on recent HPC computing nodes while component models ease the separation of concerns of application and thus improve their modularity and adaptability. We have designed a prototype version of the COMET programming model combining concepts of task-based and component models, and a preliminary version of the COMET runtime built on top of StarPU and L2C. Evaluations of the approach have been conducted on a real-world use-case analysis of a sub-part of the production application GYSELA. Results show that the approach is feasible and that it enables easy composition of independent software codes without introducing overheads. Performance results are equivalent to those obtained with a plain OpenMP based implementation. Part of this work is described in [38].

### *6.6.2. OpenMP Scheduling Heuristic for NUMA Architecture*

The recent addition of data dependencies to the OpenMP 4.0 standard provides the application programmer with a more flexible way of synchronizing tasks. Using such an approach allows both the compiler and the runtime system to know exactly which data are read or written by a given task, and how these data will be used through the program lifetime. Data placement and task scheduling strategies have a significant impact on performances when considering NUMA architectures. While numerous papers focus on these topics, none of them has made extensive use of the information available through dependencies. One can use this information to modify the behavior of the application at several levels: during initialization to control data placement and during the application execution to dynamically control both the task placement and the tasks stealing strategy, depending on the topology. This paper [26] introduces several heuristics for these strategies and their implementations in our OpenMP runtime XKAAPI. We also evaluate their performances on linear algebra applications executed on a 192-core NUMA machine, reporting noticeable performance improvement when considering both the architecture topology and the tasks data dependencies. We finally compare them to strategies presented previously by related works.

### *6.6.3. Extending OpenMP with Affinity Clause: Design and Implementation*

OpenMP 4.0 introduced dependent tasks, which give the programmer a way to express fine grain parallelism. Using appropriate OS support (such as NUMA libraries), the runtime can rely on the information in the depend clause to dynamically map the tasks to the architecture topology. Controlling data locality is one of the key factors to reach a high level of performance when targeting NUMA architectures. On this topic, OpenMP does not provide a lot of flexibility to the programmer yet, which lets the runtime decide where a task should be executed. In [27], we present a class of applications which would benefit from having such a control and flexibility over tasks and data placement. We also propose our own interpretation of the new affinity clause for the task directive, which is being discussed by the OpenMP Architecture Review Board. This clause enables the programmer to give hints to the runtime about tasks placement during the program execution, which can be used to control the data mapping on the architecture. In our proposal, the programmer can express affinity between a task and the following resources: a thread, a NUMA node, and a data. We then present an implementation of this proposal in the Clang-3.8 compiler, and an implementation of the corresponding extensions in our OpenMP runtime LIBKOMP. Finally , we present a preliminary evaluation of this work running two task-based OpenMP kernels on a 192-core NUMA architecture, that shows noticeable improvements both in terms of performance and scalability.

### 6.6.4. Support of High Task Throughput for Complex OpenMP Application

In [4], we present block algorithms and their implementation for the parallelization of sub-cubic Gaussian elimination on shared memory architectures using OpenMP standard. Contrarily to the classical cubic algorithms in parallel numerical linear algebra, we focus here on recursive algorithms and coarse grain parallelization. Indeed, sub-cubic matrix arithmetic can only be achieved through recursive algorithms making coarse grain block algorithms perform more efficiently than fine grain ones. This work is motivated by the design and implementation of dense linear algebra over a finite field, where fast matrix multiplication is used extensively and where costly modular reductions also advocate for coarse grain block decomposition. We incrementally build efficient kernels, for matrix multiplication first, then triangular system solving, on top of which a recursive PLUQ decomposition algorithm is built. We study the parallelization of these kernels using several algorithmic variants: either iterative or recursive and using different splitting strategies. Experiments show that recursive adaptive methods for matrix multiplication, hybrid recursive-iterative methods for triangular system solve and tile recursive versions of the PLUQ decomposition, together with various data mapping policies, provide the best performance on a 32 cores NUMA architecture. Overall, we show that the overhead of modular reductions is more than compensated by the fast linear algebra algorithms and that exact dense linear algebra matches the performance of full rank reference numerical software even in the presence of rank deficiencies.

## 6.7. Security for Virtualization and Clouds

**Participants:** Eddy Caron, Arnaud Lefray.

### 6.7.1. Secured Systems in Clouds with Model-Driven Orchestration

As its complexity grows, securing a system is harder than it looks. Even with efficient security mechanisms, their configuration remains a complex task. Indeed, the current practice is the hand-made configuration of these mechanisms to protect systems about which we generally lack information. Cloud computing brings its share of new security concerns but it may also be considered as leverage to overcome these issues. In [13] we adressed the key challenge of achieving global security of Cloud systems and advocate for a new approach: Model-Driven Orchestration. We have designed an implementation of this new approach called Security-Aware Models for Clouds. With this approach an industrial use-case has been deployed and secured using the Sam4C software.

## 6.8. Large Scale Cloud Deployment

**Participants:** Eddy Caron, Marcos Dias de Assunção, Christian Perez, Pedro de Souza Bento Da Silva.

### 6.8.1. Efficient Heuristics for Placing Large-Scale Distributed Applications on Multiple Clouds

With the fast growth of the demand for Cloud computing services, the Cloud has become a very popular platform to develop distributed applications. Features that in the past were available only to big corporations, like fast scalability, availability, and reliability, are now accessible to any customer, including individuals and small companies, thanks to Cloud computing. In order to place an application, a designer must choose among VM types, from private and public cloud providers, those that are capable of hosting her application or its parts using as criteria application requirements, VM prices, and VM resources. This procedure becomes more complicated when the objective is to place large component based applications on multiple clouds. In this case, the number of possible configurations explodes making necessary the automation of the placement. In this context, scalability has a central role since the placement problem is a generalization of the NP-Hard multi-dimensional bin packing problem.

In this work [22], we propose efficient greedy heuristics based on first fit decreasing and best fit algorithms, which are capable of computing near optimal solutions for very large applications, with the objective of minimizing costs and meeting application performance requirements. Through a meticulous evaluation, we show that the greedy heuristics took a few seconds to calculate near optimal solutions to placements that would require hours or even days when calculated using state of the art solutions, namely exact algorithms or meta-heuristics.

### 6.8.2. Multi-Criteria Malleable Task Management for Hybrid-Cloud Platforms

The use of large distributed computing infrastructure is a mean to address the ever increasing resource demands of scientific and commercial applications. The scale of current large-scale computing infrastructures and their heterogeneity make scheduling applications an increasingly complex task. Cloud computing minimises the heterogeneity by using virtualization mechanisms, but poses new challenges to middleware developers, such as the management of virtualization, elasticity and economic models. In this context, we proposed algorithms for efficient scheduling and execution of malleable computing tasks with high granularity while taking into account multiple optimisation criteria such as resource cost and computation time. We focused on hybrid platforms that comprise both clusters and cloud providers. In [12] we defined and formalized the main aspects of the problem, introduced the difference between local and global scheduling algorithms and evaluate their efficiency using discrete-event simulation.

## 6.9. Workflow management on Cloud environment

**Participants:** Daniel Balouek-Thomert, Eddy Caron, Laurent Lefevre.

### 6.9.1. Multi-objective workflow placements in Clouds

The recent rapid expansion of Cloud computing facilities triggers an attendant challenge to facility providers and users for methods for optimal placement of workflows on distributed resources, under the often-contradictory impulses of minimizing makespan, energy consumption, and other metrics. Evolutionary Optimization techniques that from theoretical principles are guaranteed to provide globally optimum solutions, are among the most powerful tools to achieve such optimal placements. Multi-Objective Evolutionary algorithms by design work upon contradictory objectives, gradually evolving across generations towards a converged Pareto front representing optimal decision variables – in this case the mapping of tasks to resources on clusters. However the computation time taken by such algorithms for convergence makes them prohibitive for real time placements because of the adverse impact on makespan. In [11], we described parallelization, on the same cluster, of a Multi-objective Differential Evolution method (NSDE-2) for optimization of workflow placement, and the attendant speedups that bring the implicit accuracy of the method into the realm of practical utility. We did experimental validation on a reallife testbed using diverse Cloud traces. The solutions under different scheduling policies demonstrate significant reduction in energy consumption with some improvement in makespan. We designed, implementation and evaluation of an energy-efficient resource management system that builds upon DIET, an open source middleware and NSDE-divisible tasks with precedence constraints. Real-life experiment of this approach on the Grid'5000 testbed demonstrates its effectiveness in a dynamic environment.

# 7. Bilateral Contracts and Grants with Industry

## 7.1. Bilateral Grants with Industry

### 7.1.1. NewGeneration-SR

We have a collaboration with the company NewGeneration-SR (http://newgeneration-sr.com/). The aim of this company is to reduce the energy impact through solutions at each layer of the energy consumption (from datacenter design and the production to usage). NewGeneration-SR improves the life cycle (design, production, recycling) in order to reduce the environmental impact of it. NewGeneration-SR was member of the Nu@ge consortium: one of five national Cloud Computing projects with "emprunts d'avenir" funding. With a CIFRE PhD student (Daniel Balouek-Thomert), we are developing models to reduce the energy consumption for the benefit of data-center.

### *7.1.2. IFPEN*

We also have a collaboration with IFPEN (http://ifpenergiesnouvelles.com/). IFPEN develops numerical codes to solve PDE with specific adaption of the preconditioning step to fit the requirement of their problems. With a PhD student (Adrien Roussel) we are studying the parallel implementation of multi-level decomposition domains on many-core architecture and GPGPU.

# 8. Partnerships and Cooperations

## 8.1. National Initiatives

### *8.1.1. PIA*

#### *8.1.1.1. PIA ELCI, Environnement Logiciel pour le Calcul Intensif, 2014-2017*

**Participants:** Hélène Coullon, Thierry Gautier, Laurent Lefevre, Christian Perez, Issam Rais, Jérôme Richard.

The ELCI PIA project is coordinated by BULL with several partners: CEA, Inria, SAFRAB, UVSQ.

This project aims to improve the support for numerical simulations and High Performance Computing (HPC) by providing a new generation software stack to control supercomputers, to improve numerical solvers, and pre- and post computing software, as well as programming and execution environment. It also aims to validate the relevance of these developments by demonstrating their capacity to deliver better scalability, resilience, modularity, abstraction, and interaction on some application use-cases. Avalon is involved in WP1 and WP3 ELCI Work Packages through the PhD of Issam Rais and the postdoc of Hélène Coullon. Laurent Lefevre is the Inria representative in the ELCI technical committee.

### *8.1.2. French National Research Agency Projects (ANR)*

#### *8.1.2.1. ANR INFRA MOEBUS, Multi-objective scheduling for large computing platforms, 4 years, ANR-13-INFR-000, 2013-2016*

**Participants:** Laurent Lefevre, Salem Harrache, Olivier Mornard, Christian Perez, Frédéric Suter.

The ever growing evolution of computing platforms leads to a highly diversified and dynamic landscape. The most significant classes of parallel and distributed systems are supercomputers, grids, clouds and large hierarchical multi-core machines. They are all characterized by an increasing complexity for managing the jobs and the resources. Such complexity stems from the various hardware characteristics and from the applications characteristics. The MOEBUS project focuses on the efficient execution of parallel applications submitted by various users and sharing resources in large-scale high-performance computing environments.

We propose to investigate new functionalities to add at low cost in actual large scale schedulers and programming standards, for a better use of the resources according to various objectives and criteria. We propose to revisit the principles of existing schedulers after studying the main factors impacted by job submissions. Then, we will propose novel efficient algorithms for optimizing the schedule for unconventional objectives like energy consumption and to design provable approximation multi-objective optimization algorithms for some relevant combinations of objectives. An important characteristic of the project is its right balance between theoretical analysis and practical implementation. The most promising ideas will lead to integration in reference systems such as SLURM and OAR as well as new features in programming standards implementations such as MPI or OpenMP.

#### *8.1.2.2. ANR INFRA SONGS, Simulation Of Next Generation Systems, 4 years, ANR-12-INFRA-11, 2012-2016*

**Participant:** Frédéric Suter.

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 manage such platforms still raises many challenges. As demonstrated by the USS SIMGRID project, 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 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 lead 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.

### *8.1.3. Inria Large Scale Initiative*

#### *8.1.3.1. C2S@Exa, Computer and Computational Sciences at Exascale, 4 years, 2013-2017*

**Participants:** Hélène Coullon, Laurent Lefevre, Christian Perez, Jérôme Richard, Thierry Gautier.

Since January 2013, the team is participating to the C2S@Exa Inria Project Lab (IPL). This national initiative aims at the development of numerical modeling methodologies that fully exploit the processing capabilities of modern massively parallel architectures in the context of a number of selected applications related to important scientific and technological challenges for the quality and the security of life in our society. At the current state of the art in technologies and methodologies, a multidisciplinary approach is required to overcome the challenges raised by the development of highly scalable numerical simulation software that can exploit computing platforms offering several hundreds of thousands of cores. Hence, the main objective of C2S@Exa is the establishment of a continuum of expertise in the computer science and numerical mathematics domains, by gathering researchers from Inria project-teams whose research and development activities are tightly linked to high performance computing issues in these domains. More precisely, this collaborative effort involves computer scientists that are experts of programming models, environments and tools for harnessing massively parallel systems, algorithmists that propose algorithms and contribute to generic libraries and core solvers in order to take benefit from all the parallelism levels with the main goal of optimal scaling on very large numbers of computing entities and, numerical mathematicians that are studying numerical schemes and scalable solvers for systems of partial differential equations in view of the simulation of very large-scale problems.

#### *8.1.3.2. DISCOVERY, DIstributed and COoperative management of Virtual Environments autonomousLY, 4 years, 2015-2019*

**Participants:** Jad Darrous, Gilles Fedak, Christian Perez.

To accommodate the ever-increasing demand for Utility Computing (UC) resources, while taking into account both energy and economical issues, the current trend consists in building larger and larger Data Centers in a few strategic locations. Although such an approach enables UC providers to cope with the actual demand while continuing to operate UC resources through centralized software system, it is far from delivering sustainable and efficient UC infrastructures for future needs.

The DISCOVERY initiative aims at exploring a new way of operating Utility Computing (UC) resources by leveraging any facilities available through the Internet in order to deliver widely distributed platforms that can better match the geographical dispersal of users as well as the ever increasing demand. Critical to the emergence of such locality-based UC (LUC) platforms is the availability of appropriate operating mechanisms. The main objective of DISCOVERY is to design, implement, demonstrate and promote the LUC Operating System (OS), a unified system in charge of turning a complex, extremely large-scale and widely distributed infrastructure into a collection of abstracted computing resources which is efficient, reliable, secure and at the same time friendly to operate and use.

To achieve this, the consortium is composed of experts in research areas such as large-scale infrastructure management systems, network and P2P algorithms. Moreover two key network operators, namely Orange and RENATER, are involved in the project.

By deploying and using such a LUC Operating System on backbones, our ultimate vision is to make possible to host/operate a large part of the Internet by its internal structure itself: A scalable set of resources delivered by any computing facilities forming the Internet, starting from the larger hubs operated by ISPs, government and academic institutions, to any idle resources that may be provided by end-users.

*8.1.3.3. HAC SPECIS, High-performance Application and Computers, Studying PErformance and Correctness In Simulation, 4 years, 2016-2020*

**Participants:** Laurent Lefevre, Frédéric Suter.

Over the last decades, both hardware and software of modern computers have become increasingly complex. Multi-core architectures comprising several accelerators (GPUs or the Intel Xeon Phi) and interconnected by high-speed networks have become mainstream in HPC. Obtaining the maximum performance of such heterogeneous machines requires to break the traditional uniform programming paradigm. To scale, application developers have to make their code as adaptive as possible and to release synchronizations as much as possible. They also have to resort to sophisticated and dynamic data management, load balancing, and scheduling strategies. This evolution has several consequences:

First, this increasing complexity and the release of synchronizations are even more error-prone than before. The resulting bugs may almost never occur at small scale but systematically occur at large scale and in a non deterministic way, which makes them particularly difficult to identify and eliminate.

Second, the dozen of software stacks and their interactions have become so complex that predicting the performance (in terms of time, resource usage, and energy) of the system as a whole is extremely difficult. Understanding and configuring such systems therefore becomes a key challenge.

These two challenges related to correctness and performance can be answered by gathering the skills from experts of formal verification, performance evaluation and high performance computing. The goal of the HAC SPECIS Inria Project Laboratory is to answer the methodological needs raised by the recent evolution of HPC architectures by allowing application and runtime developers to study such systems both from the correctness and performance point of view.

## 8.2. European Initiatives

### *8.2.1. FP7 & H2020 Projects*

*8.2.1.1. PaaSage*

**Participants:** Pedro de Souza Bento Da Silva, Matthieu Imbert, Christian Perez.

Title: PaaSage: Model-based Cloud Platform Upperware

Type: Seventh Framework Programme

Instrument: Collaborative project

Duration: October 2012 - September 2016 (48 months)

Coordinator: Pierre Guisset (GEIE ERCIM)

Others partners: SINTEF, STFC, HLRS, University of Stuttgart, Inria, CETIC, FORTH, be.wan, EVRY, SysFera, Flexiant, Lufthansa Systems, AG GWDG, Automotive Simulation Center Stuttgart e.V.

See also: http://paasage.eu

Abstract: PaaSage will deliver an open and integrated platform, to support both deployment and design of Cloud applications, together with an accompanying methodology that allows model-based development, configuration, optimization, and deployment of existing and new applications independently of the existing underlying Cloud infrastructures. Specifically it will deliver an IDE (Integrated Development Environment) incorporating modules for design time and execution time optimizations of applications specified in the CLOUD Modeling Language (CLOUD ML), execution-level mappers and interfaces and a metadata database.

### 8.2.2. Collaborations in European Programs, Except FP7 & H2020

#### 8.2.2.1. CHIST-ERA STAR

**Participants:** Radu Carpa, Marcos Dias de Assunção, Olivier Glück, Laurent Lefevre.

Title: SwiTching And tRansmission project

Type: CHIST-ERA (European Coordinated Research on Long-term Challenges in Information and Communication Sciences & Technologies ERA-Net)

Duration: 2013-2016

Coordinator: Jaafar Elmirghani (University of Leeds - UK)

Others partners: Inria ,University of Cambridge (UK), University of Leeds (UK), AGH University of Science and Technology Department of Telecommunications (Poland)

See also: http://www.chistera.eu/projects/star

Abstract: The Internet power consumption has continued to increase over the last decade as a result of a bandwidth growth of at least 50 to 100 times. Further bandwidth growth between 40% and 300% is predicted in the next 3 years as a result of the growing popularity of bandwidth intensive applications. Energy efficiency is therefore increasingly becoming a key priority for ICT organizations given the obvious ecological and economic drivers. In this project we adopt the GreenTouch energy saving target of a factor of a 100 for Core Switching and Routing and believe this ambitious target is achievable should the research in this proposal proven successful. A key observation in core networks is that most of the power is consumed in the IP layer while optical transmission and optical switching are power efficient in comparison, hence the inspiration for this project. Initial studies by the applicants show that physical topology choices in networks have the potential to significantly reduce the power consumption, however network optimization and the consideration of traffic and the opportunities afforded by large, low power photonic switch architectures will lead to further power savings. Networks are typically over provisioned at present to maintain quality of service. We will study optimum resource allocation to reduce the over-provisioning factor while maintaining the quality of service. Protection is currently provided in networks through the allocation of redundant paths and resources, and for full protection there is a protection route for every working route. Avalon is contributing to STAR in terms of software network protocols and services optimizations which will be combined with more efficient photonic switches in order to obtain a factor of 100 power saving in core networks. Avalon has put in place and deployed several experimental hardware (NetFPGA, low power processors, high performance servers) and software (SDN) platforms in order to validate the various energy efficient services.

#### 8.2.2.2. COST IC1305 : Nesus

**Participants:** Marcos Dias de Assunção, Laurent Lefevre, Violaine Villebonnet.

Program: COST

Project acronym: IC1305

Project title: Network for Sustainable Ultrascale Computing (NESUS)

Duration: 2014-2019

Coordinator: Jesus Carretero (Univ. Madrid)

Abstract: Ultrascale systems are envisioned as large-scale complex systems joining parallel and distributed computing systems that will be two to three orders of magnitude larger that today's systems. The EU is already funding large scale computing systems research, but it is not coordinated across researchers, leading to duplications and inefficiencies. The goal of the NESUS Action is to establish an open European research network targeting sustainable solutions for ultrascale computing aiming at cross fertilization among HPC, large scale distributed systems, and big data management. The network will contribute to glue disparate researchers working across different areas and provide a meeting ground for researchers in these separate areas to exchange ideas, to identify synergies, and to pursue common activities in research topics such as sustainable software solutions (applications and system software stack), data management, energy efficiency, and resilience. In Nesus, Laurent Lefevre is co-chairing the Working on Energy Efficiency (WG5).

# 8.3. International Initiatives

## 8.3.1. Inria International Labs

### 8.3.1.1. Joint Laboratory for Extreme Scale Computing (JLESC) (2014-2018)

**Participants:** Hélène Coullon, Gilles Fedak, Thierry Gautier, Vincent Lanore, Christian Perez, Jérôme Richard.

Partners: NCSA (US), ANL (US), Inria (FR), Jülich Supercomputing Centre (DE), BSC (SP), Riken (JP).
The purpose of the Joint Laboratory for Extreme Scale Computing (JLESC) is to be an international, virtual organization whose goal is to enhance the ability of member organizations and investigators to make the bridge between Petascale and Extreme computing. The founding partners of the JLESC are Inria and UIUC. Further members are ANL, BSC, JSC and RIKEN-AICS.

JLESC involves computer scientists, engineers and scientists from other disciplines as well as from industry, to ensure that the research facilitated by the Laboratory addresses science and engineering's most critical needs and takes advantage of the continuing evolution of computing technologies.

### 8.3.1.2. Associate Team DALHIS – Data Analysis on Large-scale Heterogeneous Infrastructures for Science (2013-2018)

**Participant:** Frédéric Suter.

Partners: EPC Myriads (Rennes, Bretagne Atlantique), Avalon (Grenoble, Rhône-Alpes), Data Science and Technology Department (LBNL,USA).

The goal of the Inria-LBL collaboration is to create a collaborative distributed software ecosystem to manage data lifecycle and enable data analytics on distributed data sets and resources. Specifically, our goal is to build a dynamic software stack that is user-friendly, scalable, energy-efficient and fault tolerant. We plan to approach the problem from two dimensions: (i) Research to determine appropriate execution environments that allow users to seamlessly execute their end-to-end dynamic data analysis workflows in various resource environments and scales while meeting energy-efficiency, performance and fault tolerance goals; (ii) Engagement in deep partnerships with scientific teams and use a mix of user research with system software R&D to address specific challenges that these communities face, and inform future research directions from acquired experience.

### 8.3.1.3. Informal International Partners

- Université Gaston Berger, Saint Louis, Sénégal. Contact: Pr. Ousmane Thiaré.
- École Centrale Mahindra, Hyderabad, India. Contact: Dr. Arya Kumar Bhattacharya.
- Center for Computing and Networking, Chinese Academy of Sciences, Beijing, China. Pr. Haiwu He.

# 8.4. International Research Visitors

## 8.4.1. Visits of International Scientists

Alberto Cabrera, University of La Laguna, Spain, Jan 2016

Damian Fernandez Cerero, University of Sevilla, Spain, Sep 2016-Dec 2016

Pr. Haiwu He, Computer Network Information Center, Chinese Academy of Sciences, Beijing, China, Nov 2016-Jan 2017

Tchimou N'Takpé, Université Nangui Abrogoua, Côte d'Ivoire, Nov 2016-Dec 2016.

### 8.4.1.1. Internships

Daniel Ciugurean, University of Cluj, Romania, Jun-Sep 2016

Joel Faubert, University of Ottawa, Canada, May-Jul 2016

# 9. Dissemination

## 9.1. Promoting Scientific Activities

### *9.1.1. Scientific Events Organisation*

*9.1.1.1. General Chair, Scientific Chair*

- Laurent Lefevre was :
  - Co-Workshop chair of Xgreen 2016 : Energy Efficiency in Large Scale Distributed Systems, Cartagena, Colombia, May 2016
  - Co-chair of the Special Session on Energy Efficient Management of Parallel Systems, Platforms, and Computations during PDP2016 : 24th Euromicro International Conference on Parallel, Distributed and Network based Processing, Heraklion, Greece, February 17-19, 2016
- Christian Perez was co-chair of the Topic 8 of Euro-Par 2016.

*9.1.1.2. Member of the Organizing Committees*

- Eddy Caron was:
  - Co-organizer of SC16 Inria Booth at Salt Lake City (Utah). November 13-18, 2016.
  - Co-organizer of SSS'2016 (The 18th International Symposium on Stabilization, Safety, and Security of Distributed Systems), Lyon, November 7-10, 2016.
  - Co-organizer of OpenStack WorkShop Lyon 2016 at ENS de Lyon. June 15, 2016.
- Hélène Coullon and Christian Perez were members of the organization team of the 5th workshop of the Joint Laboratory for Extreme Scale Computing, Lyon, 27-29 June 2016.
- Laurent Lefevre was :
  - Co-organizer of SC16 Inria Booth at Salt Lake City (Utah). November 13-18, 2016.
  - Co-organizer of the Ecoinfo conference on Eco-design of hardware", Grenoble, September 27, 2016
  - Co-organizer of the Atelier Compas 2016 : Economiser l'énergie dans les systèmes distribués, July 5, 2016 with Anne-Cécile Orgerie (IRISA, Rennes)
  - Co-organizer of the E3-RSD school on Energy Efficiency in Networks and Distributed Systems, Dinard, May 23-27, 2016 with Anne-Cécile Orgerie (IRISA, Rennes)
  - Tutorial Chair in CCGrid 2016: the 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, Cartagena, Colombia, May 2016
- Frédéric Suter was the co-organizer of the 6th edition of the SimGrid Users Days, Fréjus, January 18-21, 2016.

### *9.1.2. Scientific Events Selection*

*9.1.2.1. Member of the Conference Program Committees*

- Eddy Caron was a member of the program committee of HCW'2016, ICACCI-2016, Xgreen 2016, CloudTech'16, CCGRID'2016
- Gilles Fedak was a member of the program committee of InterCloud'2016, CLoudCom'2016, Cluster 2016, Compass 2016, HPDC 2016, ICCCN'2016, ICPADS'2016.
- Olivier Glück was a member of the program committee of PDP 2016 (24th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing) and the Xgreen2016 workshop of CCGrid 2016.

- Christian Perez was a member of the Workshop Committee at the 2016 ACM/IEEE International Conference on High Performance Computing, Networking, Storage and Analysis (SC16). He was a member of the program committee of the Grid'5000 winter School 2016. He served in the committee to select Outstanding Paper Award of HPCS 2016.
- Frédéric Suter was a member of the program committee of Vecpar 2016, ICCS 2016, ComPas 2016 and EuroMPI 2016

### *9.1.3. Journal*

*9.1.3.1. Member of the Editorial Boards*

- G. Fedak: Co-editor of Elsevier Journal of Cluster Computing

### *9.1.4. Invited Talks*

Laurent Lefevre has been invited to give the following talks :

- "Reducing environmental impacts of ICT : challenges and solutions", Invited talk in "Agence de l'Eau Rhône Méditerranée et Corse", Lyon, December 15, 2016
- "Ecological digital world, is it possible ? / Numérique écologique, est ce possible ?", Invited Keynote for Université Ouverte de Lyon, Lyon, France, November 29, 2016
- "Grid'5000: French Nationwide Testbed for Research in Distributed Computing", Laurent Lefevre, BOF on Experimental Infrastructure and Methodology for HPC Cloud, Supercomputing 2016 Conference, Salt Lake City, USA, November 15, 2016
- "GreenFactory : orchestrating power capabilities and leverages at large scale for energy efficient infrastructures", Laurent Lefevre, CCDSC2016: Workshop on Clusters, Clouds, and Data for Scientific Computing, Dareize, France, October 3-6, 2016
- "Full life cycle in sustainable ICT for reaching energy reduction, energy efficiency and energy proportionality", Laurent Lefevre, Invited keynote talk, IEEE Tenth International Conference on Research Challenges in Information Science (RCIS2016) & Inforsid conference, Grenoble France, June 3, 2016

Gilles Fedak has been invited to give the following talks:

- “iEx.ec: Blockchain-based Cloud Computing”, Inria Blockchain Day, Paris, December 12, 2016
- “Active Data: Large scale data management based on data life cycle”, Journée Big Data de la Fédération Informatique de Lyon, December 9, 2016
- “Active Data, managing Data Life Cycle on heterogeneous systems and infrastructures”, Workshop Data preservation (PREDON) of the MADICS research group on Big Data, Paris, CNRS, December 2, 2016
- “Blockchain applications are coming, what's next” Panel at the workshop ParisInvest, HEC, December 1st, 2016
- “iEx.ec: Blockchain-based Distributed Cloud Computing”, Chaintech Meetup, Paris, November 29, 2016
- “iEx.ec: un Cloud distribué basé sur la blockchain”, CES-3 Innovation Forum, Paris, 29 Septembre 2016
- “iEx.ec: un Cloud distribué basé sur la blockchain”, EDF research Lab, Saclay, 28 Septembre 2016
- “iEx.ec: Blockchain-based Fully Distributed Cloud Computing”,CAS Shanghai University, Shanghai, September 23, 2016
- “iEx.ec: Fully Distributed Cloud Computing thanks to the Ethereum Smart Contracts”, Ethereum Developper Conference, Shanghai, September 21, 2016
- “iEx.ec: Blockchain-based Fully Distributed Cloud Computing”, CNIC Chinese Academy of Sciences, Beijing, September 18, 2016
- "Relocaliser les data-centers dans la ville intelligente grâce à la blockchain" Smart Building Alliance, Lyon, France September 15, 2016

Frédéric Suter has been invited to give the following talk:

- "Modeling Distributed Platforms from Application Traces for Realistic File Transfer Simulation", Information Sciences Institute, University of Southern California, December 21, 2016.

### 9.1.5. Leadership within the Scientific Community

- Laurent Lefevre is Animator and chair of the transversal action on "Energy" of the French GDR RSD ("Réseaux et Systèmes Distribués")
- Christian Perez is co-leader of the pole Distributed Systems of the French GDR RSD ("Réseaux et Systèmes Distribués").

### 9.1.6. Scientific Expertise

- Eddy Caron reviewed a project for the French National Research Agency (ANR).
- Olivier Glück is member of the CNU (Conseil National des Universités) section 27 (Computer Science). He participated to the 2016 "Qualifications" session.
- Olivier Glück evaluated a Ph.D. CIFRE proposal (industrial funded PhD student).
- Christian Perez reviewed a project for the French National Research Agency (ANR).

### 9.1.7. Research Administration

- Eddy Caron is a member of the Inria's CDT (Commission de Développement Technologique) from Inria Rhone-Alpes center.
- Eddy Caron is the corresponding researcher "valorisation/transfert" in the Labex MILyon for the LIP laboratory.
- Olivier Glück is member of the "Conseil Académique" of Lyon 1 and Lyon University. He is also member of the "Conseil de la Faculté des Sciences et Technologies" of Lyon 1 University.
- Christian Perez is a member of the committee selection of the project call from the Rhône-Alpes region ARC6 (France). He is also a member of the Inria Grenoble Rhône-Alpes Strategic Orientation Committee.

## 9.2. Teaching - Supervision - Juries

### 9.2.1. Teaching

Licence: Yves Caniou, Algorithmique programmation impérative, initiation, 10h, niveaux L1, Université Lyon 1, France

Licence: Yves Caniou, Système d'Exploitation, 35h, niveaux L3, Université Lyon 1, France

Licence: Yves Caniou, Système d'Exploitation, 35h, niveaux L2, Université Lyon 1, France

Licence: Yves Caniou, Programmation Concurrente, 33h, niveaux L3, Université Lyon 1, France

Licence: Yves Caniou, Réseaux, 36h, niveaux L3, Université Lyon 1, France

Master: Yves Caniou, Sécurité, 34h, niveaux M2, Université Lyon 1, France

Master: Yves Caniou, Projet, Bibliographie, Certification, 15h, niveaux M2, Université Lyon 1, France

Master: Yves Caniou, Sécurité, 20h, niveau M2, IGA Casablanca, Maroc

Master: Yves Caniou, Suivi d'étudiants (apprentissage, stage), 25h, niveaux M1 et M2, Université Lyon 1, France

Master: Yves Caniou, Responsible of professional Master SRIV (Systèmes Réseaux et Infrastructures Virtuelles), 30h, Université Lyon 1, France

Licence: Eddy Caron, Projet 1, 48h, L3, Ens de Lyon. France.

Licence: Eddy Caron, Architecture, Système et Réseaux, 48h, L3, Ens de Lyon. France.

Master: Eddy Caron, Projet Intégré, 42h, M1, Ens de Lyon. France.

Master: Eddy Caron, Système distribués, 30h, M1, Ens de Lyon. France.

Master: Eddy Caron, Distributed Computing: Models and Challenges, 8h, M2, Ens de Lyon. France.

Licence: Olivier Glück, Licence pedagogical advisor, 30h, niveaux L1, L2, L3, Université Lyon 1, France.

Licence: Olivier Glück, Initiation Réseaux, 2x9h, niveau L2, Université Lyon 1, France.

Licence: Olivier Glück, Réseaux, 2x70h, niveau L3, Université Lyon 1, France.

Master: Olivier Glück, Responsible of professional Master SIR (Systèmes Informatiques et Réseaux) located at IGA Casablanca, 20h, niveau M2, IGA Casablanca, Maroc

Master: Olivier Glück, Réseaux par la pratique, 20h, niveau M1, Université Lyon 1, France.

Master: Olivier Glück, Services et Protocoles Applicatifs sur Internet, 40h, niveau M2, Université Lyon 1, France.

Master: Olivier Glück, Services et Protocoles Applicatifs sur Internet, 24h, niveau M2, IGA Casablanca, Maroc

Master: Olivier Glück, Administration des Systèmes et des Réseaux, 16h, niveau M2, Université Lyon 1, France.

Master: Jean-Patrick Gelas, Programmation embarquée et mobile des objets, 17.5h, niveau M1, Université Lyon 1, France.

Master: Jean-Patrick Gelas, Introduction au Cloud Computing, 24h, niveau M2 (CCI), Université Lyon 1, France.

Master: Jean-Patrick Gelas, Système d'exploitation, 30h, niveau M2 (CCI), Université Lyon 1, France.

Master: Jean-Patrick Gelas, Projet en Informatique en Anglais, 10h, niveau M2 (CCI), Université Lyon 1, France.

Master: Jean-Patrick Gelas, Réseaux Avancés, 27h, niveau M2 (CCI), Université Lyon 1, France.

Master: Jean-Patrick Gelas, Sécurité et Admin des infra réseaux, 37.5h, niveau M2 (CCI), Université Lyon 1, France.

Master: Jean-Patrick Gelas, Technologies embarquées, 15h, niveau M2 (Image), Université Lyon 1, France.

Master: Jean-Patrick Gelas, Routage (BGP), Routeurs et IPv6, 16.5h, niveau M2, Université Lyon 1, France.

Master: Jean-Patrick Gelas, Systèmes embarqués (GNU/Linux, Android, ARM, Arduino), 39h, niveau M2, Université Lyon 1, France.

Master: Jean-Patrick Gelas, Analyse de performance, 3h, niveau M2 (TIW), Université Lyon 1, France.

Master: Jean-Patrick Gelas, Cloud Computing, 16h, niveau M2 (TIW), Université Lyon 1, France.

Master Informatique: Laurent Lefevre, "Parallelism", Université Claude Bernard, France. (18h), M1

Master Systèmes Informatique et Réseaux: Laurent Lefevre, "Advanced Networks", IGA Casablanca, Maroc (20h), M2

### 9.2.2. Supervision

PhD: Daniel Balouek-Thomert, *Ordonnancement et éco-efficacité dans les environnements virtualisés*, 09/2013, Eddy Caron (dir), Laurent Lefevre (co-dir), Gilles Cieza (NewGen society, co-dir), defended December 5th, 2016

PhD: Violaine Villebonnet, *Proportionnalité énergétique dans les systèmes distribués à grande échelle*, 9/2013, Laurent Lefevre (dir), Jean-Marc Pierson (IRIT, Toulouse, co-dir), defended December 6th, 2016

PhD in progress: Radu Carpa, *Efficacité énergétique des échanges de données dans une fédération d'infrastructures distribuées à grande échelle*, 10/2014, Laurent Lefèvre (dir), Olivier Glück (co-dir).

PhD in progress: Hadrien Croubois, *Étude et conception d'un système de gestion de workflow autonomique conçu pour l'animation 3D* , 10/2015, Eddy Caron.

PhD in progress: Pedro Silva, *Application model and co-scheduling algorithm for dynamic and evolutive data-intensive application*, 10/2014, Christian Perez (dir), Frédéric Desprez (co-dir).

PhD in progress: Issam Rais, *Multi criteria scheduling for exascale infrastructures*, 10/2014, Laurent Lefevre (dir), Anne Benoit (Roma Team, LIP, ENS Lyon, co-dir) and Anne-Cécile Orgerie (CNRS, Myriads team, Irisa Rennes, co-dir)

PhD in progress: Jérôme Richard, *Conception of a software component model with task scheduling for many-core based parallel architecture, application to the Gysela5D code*, 11/2014, Christian Perez (dir), Julien Bigot (CEA, co-dir).

PhD in progress: Alexandre Veith : *Elastic Mechanisms for Big-Data Stream Analytics*, Labex MiLyon, Laurent Lefevre (dir), Marcos Dias de Assuncao (co-dir) (2016-2019)

PhD in progress: Valentin Lorentz : *Energy traceability of data*, Gilles Fedak (dir), Laurent Lefevre (co-dir) (2016-2019)

PhD in progress: Hayri Acar, *Towards a green and sustainable software*, Parisa Ghodous (dir), Gulfem Alptekin (co-dir), Jean-Patrick Gelas (co-dir) (2014-2017)

PhD in progress: Jad Darrous : *Geo-distributed storage for distributed Cloud*, Gilles Fedak (dir), Shadi Ibrahim (co-dir, Kerdata team, Inria Rennes) (2016-2019)

PhD in progress: Anchen Chai: *Simulation of the Distributed Execution of a Medical Imaging Simulator*, Hugues Benoit-Cattin (co-dir, CREATIS, INSA Lyon), Frédéric Suter (co-dir).

PhD in progress: Aurélie Kong-Win-Chang: *Techniques de résilience pour l'ordonnancement de workflows sur plates-formes décentralisées (cloud computing) avec contraintes de sécurité*, Yves Robert (dir, ROMA, ÉNS-Lyon), Eddy Caron (co-dir) et Yves Caniou (co-dir).

### *9.2.3. Juries*

Eddy Caron has been member of the following PhD jury:

- Mohamed Hamza Kaaouachi: "Une approche distribuée pour les problèmes de couverture dans les systèmes hautement dynamiques.", January 2016, Reviewer, UPMC.
- Divya Gupta: "Towards Performance and Dependability Benchmarking of Distributed Fault Tolerance Protocols", March 2016, Reviewer, Université Grenoble Alpes.
- Bassirou Gueye: " Services auto-adaptatifs pour les grilles pair-a`-pair", May 2016, Reviewer, Université de Reims Champagne-Ardenne.

Gilles Fedak has been member of the following PhD jury:

- Walid Saad: "Gestion de données pour le calcul scientifique dans les environnements grilles et cloud", October 2016, Reviewer, University of Sfax, Tunisia

Laurent Lefevre has been member of the following PhD juries:

- Chakadit Thaenchaikun : "Energy efficiency in wired networks : traffic engineering and switching off", University of Toulouse, November 2016, Reviewer
- Pablo Llopis Sanmillan : "Enhancing the programmability and energy efficiency of storage in HPC and virtualized environments", Universidad Carlos III de Madrid, Spain, July 2016, Jury member - president.
- Sareh Fotuhi Piraghaj : "Energy-Efficient Management of Resources in Enterprise and Container-based Clouds", University of Melbourne, Australia, June 2016, Reviewer
- Mohammed Hussein : "Energy Efficiency in LEO Satellite and Terrestrial Wired Environments", University of Toulouse, June 2016, Reviewer
- Fouad Hanna, : "Etude et développement du nouvel algorithme distribué de consensus FLC", University of Franche-Comté, Besancon, February 2016, Reviewer

Christian Perez has been member of the following HdR and PhD juries:

- Fabrice Huet, HdR: "From HPC to Big Data: Models and Tools for Large Scale Middleware", University of Nice Sophia-Antipolis, France., Feb 2016, Reviewer.
- Tomasz Buchert, PhD: "Managing large-scale, distributed systems research experiments with control-flows", University of Lorraine, France., Jan. 2016, jury member.
- Naweiluo Zhou, PhD: "Autonomic Thread Parallelism and Mapping Control for Software Transactional Memory", University of Grenoble, France., Oct 2016, president.

## 9.3. Popularization

- Laurent Lefevre has been :
  - Interviewed in "Interception" Radio Magazine on "To click, is to pollute" ("Cliquer c'est polluer"), France Inter Radio, 6 November 2016
  - Interviewed with Françoise Berthoud, "Télévision : Changement de norme = grand gâchis écologique - Site Reporterre.net, - 5 April 2016
- Gilles Fedak has been interviewed in Bitcoin.fr, November 2016

# 10. Bibliography

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[1] D. Balouek-Thomert. *Scheduling on Clouds considering energy consumption and performance trade-offs: from modelization to industrial applications* , Ecole Normale Supérieure de Lyon - ENS LYON, December 2016, https://hal.inria.fr/tel-01436822

[2] V. Villebonnet. *Scheduling and Dynamic Provisioning for Energy Proportional Heterogeneous Infrastructures*, Université de Lyon, December 2016, https://tel.archives-ouvertes.fr/tel-01419440

### Articles in International Peer-Reviewed Journals

[3] E. Caron, A. K. Datta, C. Tedeschi, F. Petit. *Self-Stabilizing Prefix Tree Based Overlay Networks*, in "International Journal of Foundations of Computer Science", 2016, vol. 27, n<sup>o</sup> 5, pp. 607–630 [*DOI :* 10.1142/S0129054116500192], http://hal.upmc.fr/hal-01347457

[4] J.-G. Dumas, T. Gautier, C. Pernet, J.-L. Roch, Z. Sultan. *Recursion based parallelization of exact dense linear algebra routines for Gaussian elimination*, in "Parallel Computing", September 2016, vol. 57, pp. 235–249 [*DOI :* 10.1016/j.parco.2015.10.003], https://hal.archives-ouvertes.fr/hal-01084238

[5] M.-D. Faye, E. Caron, O. Thiare. *A self-stabilizing algorithm for a hierarchical middleware self-adaptive deployment : specification, proof, simulations*, in "Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées", 2016, vol. 25, n<sup>o</sup> Special Issue, CNRIA 2015, pp. 1-20, https://hal.archives-ouvertes.fr/hal-01311153

[6] S. Lambert, P. Ananth, P. Vetter, K.-L. Lee, J. Li, X. Yin, H. Chow, J.-P. Gelas, L. Lefèvre, D. Chiaroni, B. Lannoo, M. Pickavet. *The road to energy efficient optical access: GreenTouch final results*, in "Journal of Optical Communications and Networking", March 2017, To appear, https://hal.inria.fr/hal-01369267

[7] M. MOCA, C. LITAN, G. C. SILAGHI, G. FEDAK. *Multi-criteria and satisfaction oriented scheduling for hybrid distributed computing infrastructures*, in "Future Generation Computer Systems", 2016, vol. 55 [*DOI :* 10.1016/J.FUTURE.2015.03.022], https://hal.inria.fr/hal-01239218

[8] A. ROUSSEL, J.-M. GRATIEN, T. GAUTIER. *Using Runtime Systems Tools to Implement Efficient Preconditioners for Heterogeneous Architectures*, in "Oil & Gas Science and Technology – Rev. IFP Energies nouvelles", November 2016, vol. 71, n$^{o}$ 6, 65 p. [*DOI :* 10.2516/OGST/2016020], https://hal-ifp.archives-ouvertes.fr/hal-01396153

[9] B. TANG, M. TANG, G. FEDAK, H. HE. *Availability/Network-aware MapReduce over the Internet*, in "Information Sciences", 2016, vol. 379, pp. 94–111, https://hal.inria.fr/hal-01426393

### Articles in Non Peer-Reviewed Journals

[10] J. CARRETERO, R. ČIEGIS, E. JEANNOT, L. LEFÈVRE, G. RÜNGER, D. TALIA, Ž. JULIUS. *HeteroPar 2014, APCIE 2014, and TASUS 2014 Special Issue*, in "Concurrency and Computation: Practice and Experience", 2016, 2 p. , https://hal.inria.fr/hal-01253278

### International Conferences with Proceedings

[11] D. BALOUEK-THOMERT, A. K. BHATTACHARYA, E. CARON, K. GADIREDDY, L. LEFÈVRE. *Parallel Differential Evolution approach for Cloud workflow placements under simultaneous optimization of multiple objectives*, in "Congress on Evolutionary Computation (IEEE CEC 2016)", Vancouver, Canada, July 2016, https://hal.inria.fr/hal-01289176

[12] E. CARON, M. DIAS DE ASSUNCAO. *Multi-Criteria Malleable Task Management for Hybrid-Cloud Platforms*, in "2nd International Conference on Cloud Computing Technologies and Applications (CloudTech'16)", Marrakech, Morocco, May 2016, https://hal.inria.fr/hal-01355682

[13] E. CARON, A. LEFRAY, J. ROUZAUD-CORNABAS. *Secured Systems in Clouds with Model-Driven Orchestration*, in "The 2nd IEEE Workshop on Security and Privacy in the Cloud (SPC 2016). In conjunction with the IEEE CNS conference", Philadelphia, United States, The 2nd IEEE Workshop on Security and Privacy in the Cloud (SPC 2016), IEEE, October 2016, https://hal.inria.fr/hal-01355681

[14] R. CARPA, M. DIAS DE ASSUNCAO, O. GLÜCK, L. LEFÈVRE, J.-C. MIGNOT. *Responsive Algorithms for Handling Load Surges and Switching Links On in Green Networks*, in "IEEE International Conference on Communications - IEEE ICC'16 - Green Communications Systems and Networks Symposium", Kuala Lumpur, Malaysia, May 2016, https://hal.inria.fr/hal-01266279

[15] M. DIAS DE ASSUNCAO, R. CARPA, O. GLÜCK, L. LEFÈVRE. *On Designing SDN Services for Energy-Aware Traffic Engineering*, in "Tridentcom2016 : 11th EAI International Conference on Testbeds and Research Infrastructures for the Development of Networks & Communities", Hangzhou, China, June 2016, https://hal.inria.fr/hal-01346596

[16] M. DIAS DE ASSUNCAO, L. LEFÈVRE, F. ROSSIGNEUX. *On the Impact of Advance Reservations for Energy-Aware Provisioning of Bare-Metal Cloud Resources*, in "CNSM 2016", Montreal, Canada, October 2016, https://hal.inria.fr/hal-01382662

[17] C. HERZOG, J.-M. PIERSON, L. LEFÈVRE. *A Multi Agent System for Understanding the Impact of Technology Transfer Offices in Green-IT*, in "PRIMA 2016 : International Conference on Principles and practice of multi-agent systems", Phuket, Thailand, August 2016, https://hal.inria.fr/hal-01355687

[18] C. HERZOG, J.-M. PIERSON, L. LEFÈVRE. *Modelling technology transfer in Green IT with Multi Agent System*, in "ICLIE'16 : The International Conference on Leadership, Innovation and Entrepreneurship", Dubai, United Arab Emirates, SPRINGER (editor), April 2016, https://hal.inria.fr/hal-01290038

[19] T. N 'TAKPÉ, F. SUTER. *Prise en compte de tâches non-prioritaires dans l'ordonnancement batch*, in "Conférence d'informatique en Parallélisme, Architecture et Système (Compas 2016)", Lorient, France, July 2016, https://hal.inria.fr/hal-01420693

[20] I. RAIS, L. LEFÈVRE, A. BENOIT, A.-C. ORGERIE. *An Analysis of the Feasibility of Energy Harvesting with Thermoelectric Generators on Petascale and Exascale Systems*, in "Workshop Optimization of Energy Efficient HPC & Distributed Systems (OPTIM 2016) - The 2016 International Conference on High Performance Computing & Simulation (HPCS 2016)", Innsbruck, Austria, Proceedings of the 2016 International Conference on High Performance Computing & Simulation (HPCS 2016), July 2016, https://hal.inria.fr/hal-01348554

[21] I. RAIS, A.-C. ORGERIE, M. QUINSON. *Impact of Shutdown Techniques for Energy-Efficient Cloud Data Centers*, in "International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP)", Granada, Spain, December 2016, https://hal.inria.fr/hal-01362530

[22] P. SILVA, C. PÉREZ, F. DESPREZ. *Efficient Heuristics for Placing Large-Scale Distributed Applications on Multiple Clouds*, in "16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid'16)", Cartagena, Colombia, 2016 16th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), May 2016 [*DOI :* 10.1109/CCGRID.2016.77], https://hal.archives-ouvertes.fr/hal-01301382

[23] V. VILLEBONNET, G. DA COSTA, L. LEFÈVRE, J.-M. PIERSON, P. STOLF. *Dynamically Building Energy Proportional Data Centers with Heterogeneous Computing Resources*, in "IEEE Cluster", Taipei, Taiwan, September 2016, short paper, https://hal.inria.fr/hal-01346600

[24] V. VILLEBONNET, G. DA COSTA, L. LEFÈVRE, J.-M. PIERSON, P. STOLF. *Energy Aware Dynamic Provisioning for Heterogeneous Data Centers*, in "SBAC-PAD : 28th International Symposium on Computer Architecture and High Performance Computing", Los Angeles, United States, October 2016, https://hal.inria.fr/hal-01355452

[25] V. VILLEBONNET, G. DA COSTA, L. LEFÈVRE, J.-M. PIERSON, P. STOLF. *Energy Proportionality in Heterogeneous Data Center Supporting Applications with Variable Load*, in "IEEE International Conference on Parallel and Distributed Systems (ICPADS)", Wuhan, China, December 2016, https://hal.inria.fr/hal-01396947

[26] P. VIROULEAU, F. BROQUEDIS, T. GAUTIER, F. RASTELLO. *Using data dependencies to improve task-based scheduling strategies on NUMA architectures*, in "Euro-Par 2016", Grenoble, France, Euro-Par 2016, August 2016, https://hal.inria.fr/hal-01338761

[27] P. VIROULEAU, A. ROUSSEL, F. BROQUEDIS, T. GAUTIER, F. RASTELLO, J.-M. GRATIEN. *Description, Implementation and Evaluation of an Affinity Clause for Task Directives*, in "IWOMP 2016", Nara, Japan, IWOMP 2016 - LLCS 9903, October 2016, https://hal.inria.fr/hal-01343442

### National Conferences with Proceedings

[28] C. HERZOG, J.-M. PIERSON, L. LEFÈVRE. *Modélisation du transfert technologique par système multi-agent : illustration à l'informatique verte*, in "JFSMA 2016 : 24 èmes journées Francophones sur les systèmes Multi-Agents", Rouen, France, October 2016, https://hal.inria.fr/hal-01369283

### Conferences without Proceedings

[29] H. ACAR, G. I. ALPTEKIN, J.-P. GELAS, P. GHODOUS. *Beyond CPU: Considering Memory Power Consumption of Software*, in "Smartgreens 2016", Rome, Italy, April 2016, https://hal.archives-ouvertes.fr/hal-01314070

[30] I. RAIS, A. BENOIT, L. LEFÈVRE, A.-C. ORGERIE. *An Analysis of the Feasibility of Energy Harvesting with Thermoelectric Generators on Petascale and Exascale Systems*, in "Conférence d'informatique en Parallélisme, Architecture et Système (COMPAS 2016)", Lorient, France, Actes de COMPAS, la Conférence d'informatique en Parallélisme, Architecture et Système, July 2016, https://hal.inria.fr/hal-01348555

[31] A. ROUSSEL. *Comparaison de moteurs exécutifs pour la parallélisation de solveurs linéaires itératifs*, in "Conférence d'informatique en Parallélisme, Architecture et Système (Compas'2016)", Lorient, France, July 2016, https://hal-ifp.archives-ouvertes.fr/hal-01343151

[32] P. VIROULEAU. *Amélioration des stratégies d'ordonnancement sur architectures NUMA à l'aidedes dépendances de données*, in "Compas 2016", Lorient, France, July 2016, https://hal.inria.fr/hal-01338750

### Research Reports

[33] H. COULLON, J. BIGOT, C. PÉREZ. *The Multi-Stencil Language: orchestrating stencils with a mesh-agnostic DSL*, Inria - Research Centre Grenoble – Rhône-Alpes, October 2016, n$^{o}$ RR-8962, 30 p. , https://hal.inria.fr/hal-01380607

[34] A. LÈBRE, J. PASTOR, F. DESPREZ. *A Ring to Rule Them All - Revising OpenStack Internals to Operate Massively Distributed Clouds: The Discovery Initiative - Where Do We Are ?*, Inria, February 2016, n$^{o}$ RT-0480, pp. 1-24, https://hal.inria.fr/hal-01320235

### Other Publications

[35] E. CARON, C. TEDESCHI. *Service Discovery system: architecture and algorithms*, January 2017, Délivrable, https://hal.inria.fr/hal-01428031

[36] E. CARON, C. TEDESCHI. *Service discovery system: Load balancing mechanisms*, January 2017, Délivrable, https://hal.inria.fr/hal-01427722

[37] A. DEGOMME, A. LEGRAND, M. S. MARKOMANOLIS, M. QUINSON, M. S. STILLWELL, F. S. SUTER. *Simulating MPI applications: the SMPI approach*, November 2016, Under review in IEEE TPDS, https://hal.inria.fr/hal-01415484

[38] C. PÉREZ, H. COULLON, J. RICHARD. *Feasibility Study of a Runtime Component-based Model Integrating Task Graph Concept on a 1D Advection Case Study*, July 2016, 21 p. , Livrable D3.2.1 du projet PIA ELCI, https://hal.inria.fr/hal-01348204

[39] Y. TANIMURA, K. SEYMOUR, E. CARON, A. AMAR, H. NAKADA, Y. TANAKA, F. DESPREZ. *Interoperability Testing for The GridRPC API Specification*, January 2017, working paper or preprint, https://hal.inria.fr/hal-01429586

## References in notes

[40] K. BERGMAN, S. BORKAR, D. CAMPBELL, W. CARLSON, W. DALLY, M. DENNEAU, P. FRANZON, W. HARROD, J. HILLER, S. KARP, S. KECKLER, D. KLEIN, R. LUCAS, M. RICHARDS, A. SCARPELLI, S. SCOTT, A. SNAVELY, T. STERLING, R. S. WILLIAMS, K. YELICK. *ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems*, in "DARPA Information Processing Techniques Office", Washington, DC, September 28 2008, 278 p.

[41] R. GE, X. FENG, S. SONG, H.-C. CHANG, D. LI, K. W. CAMERON. *PowerPack: Energy Profiling and Analysis of High-Performance Systems and Applications*, in "IEEE Trans. Parallel Distrib. Syst.", May 2010, vol. 21, n$^{o}$ 5, pp. 658–671 [*DOI :* 10.1109/TPDS.2009.76], http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4906989

[42] A. GEIST, S. DOSANJH. *IESP Exascale Challenge: Co-Design of Architectures and Algorithms*, in "Int. J. High Perform. Comput. Appl.", November 2009, vol. 23, n$^{o}$ 4, pp. 401–402, http://dx.doi.org/10.1177/1094342009347766

[43] W. GROPP, S. HUSS-LEDERMAN, A. LUMSDAINE, E. LUSK, B. NITZBERG, W. SAPHIR, M. SNIR. *MPI: The Complete Reference – The MPI-2 Extensions*, 2, The MIT Press, September 1998, vol. 2, ISBN 0-262-57123-4

[44] H. KIMURA, T. IMADA, M. SATO. *Runtime Energy Adaptation with Low-Impact Instrumented Code in a Power-Scalable Cluster System*, in "Proceedings of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing", Washington, DC, USA, CCGRID '10, IEEE Computer Society, 2010, pp. 378–387

[45] G. MADEC. *NEMO ocean engine*, Institut Pierre-Simon Laplace (IPSL), France, 2008, n$^{o}$ 27, ISSN No 1288-1619

[46] OPENACC. *The OpenACC Application Programming Interface*, November 2011, Version 1.0, http://www.openacc-standard.org

[47] OPENMP ARCHITECTURE REVIEW BOARD. *OpenMP Application Program Interface*, July 2011, Version 3.1, http://www.openmp.org

[48] T. RISTENPART, E. TROMER, H. SHACHAM, S. SAVAGE. *Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-Party Compute Clouds*, in "Proceedings of the 16th ACM conference on Computer and communications security", New York, NY, USA, CCS '09, ACM, 2009, pp. 199–212, http://doi.acm.org/10.1145/1653662.1653687

[49] B. ROUNTREE, D. K. LOWNENTHAL, B. R. DE SUPINSKI, M. SCHULZ, V. W. FREEH, T. BLETSCH. *Adagio: Making DVS Practical for Complex HPC Applications*, in "Proceedings of the 23rd international conference on Supercomputing", New York, NY, USA, ICS '09, ACM, 2009, pp. 460–469

[50] C. SZYPERSKI. *Component Software - Beyond Object-Oriented Programming*, 2, Addison-Wesley / ACM Press, 2002, 608 p.

[51] S. VALCKE. *The OASIS3 coupler: a European climate modelling community software*, in "Geoscientific Model Development", 2013, vol. 6, pp. 373-388, doi:10.5194/gmd-6-373-2013