INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team reso

Optimized protocols and software for high performance networks

Grenoble - Rhône-Alpes

Theme : Networks and Telecommunications

Table of contents

1. Team

Research Scientist

Pascale Vicat-Blanc Primet [Team leader, Research Director (DR2) Inria, HdR]
Paulo Gonçalves [Research Associate (CR1) Inria]
Laurent Lefèvre [Research Associate (CR1) Inria]

Faculty Member

Jean-Patrick Gelas [Maître de conférences]
Olivier Glück [Maître de conférences]
Isabelle Guérin-Lassous [Professeur, HdR]
Thomas Begin [Maître de conférences]

Technical Staff

Jean-Christophe Mignot [Research Engineer CNRS – 40%]
Aurélien Cedeyn [ENS Engineer – Grid5000 until Sep. 2009]
Matthieu Imbert [Research Engineer INRIA – 40%]
Damien Ancelin [Expert Engineer INRIA -FP6 EC-GIN until march 2009]
Olivier Mornard [Expert Engineer INRIA– ANR HIPCAL]
Philippe Martinez [Expert Engineer INRIA– ANR HIPCAL since 1/10/08]
Abderhaman Cheniour [Expert Engineer INRIA – FP7 Autonomic Internet]
Oana Goga [Associated Engineer INRIA– ADT Aladdin until Sept. 2009]
Armel Soro [Associated Engineer INRIA– ADT Aladdin since Dec. 2009]
Augustin Ragon [Expert Engineer INRIA – OGF-EUROPE]

PhD Student

Lucas Nussbaum [ATER - 2008/2009]
Romaric Guillier [PhD student, INRIA - CARRIOCAS - 2006/2009]
Ludovic Hablot [PhD student, ENS, MENRT - 2006/2009]
Patrick Loiseau [PhD student, ENS , MENRT- 2006/2009]
Doreid Ammar [PhD student, INRIA , Bell Labs- 2009/2012]
Marina Sokol [PhD student, INRIA , Bell Labs- 2009/2012]
Sébastien Soudan [PhD student, ENS, MENRT - 2006/2009]
Rémi Vanier [PhD student, INRIA - AEOLUS 2006/2009]
Dinil Mon Divakaran [PhD student, INRIA-Bell Labs - 2007/2010]
Pierre-Solen Guichard [PhD student, INRIA-Bell Labs - 2008/2011]
Anne-Cécile Orgerie [PhD student - ENS, MENRT - 2008/2011]
Fabienne Anhalt [PhD student, INRIA - 2008/2011]
Guilherme Koslovski [PhD student, INRIA - 2008/2011]

Post-Doctoral Fellow

Marcos Dias de Assuncao [Postdoc INRIA - GREEN-NET since 15/9/2009]
Manoj Dahal [Postdoc INRIA - CARRIOCAS contract Nov. 2008 / Oct. 2009]
Olivier Grémillet [Postdoc INRIA Bell Labs Dec. 2008 - Dec. 2009]
Emanouelli Draminitos [Postdoc INRIA AEOLUS contract Nov. 2008 / Oct. 2009]
Ibrahim Mouhamad [Postdoc INRIA - Jan. 2009 / Sept. 2009]

Visiting Scientist

Alejandro Fernandez-Montes Gonzalez [Phd student from University of Sevilla, Spain with RESO in September-November 2009]

Administrative Assistant

Sylvie Boyer [Secretary (SAR) INRIA – 10%]
Caroline Suter [Secretary ENS – 50%]

2. Overall Objectives

2.1. Project-team presentation overview

The RESO team belongs to the “Laboratoire de l’Informatique du Parallélisme” (LIP) - Unité Mixte de Recherche (UMR) CNRS-INRIA-ENS with Université Claude Bernard of Lyon. It consists of twenty members in average, including six permanent researchers and teaching researchers. The research activities of the RESO project fits the "communicating" scientific priority of the INRIA’s strategic plan 2008-2012. In this direction, RESO is focusing on communication software, services and protocols in the context of high speed networks and applying its results to the domain of high demanding applications and Future Internet.

The RESO approach relies on the theoretical and practical analysis of limitations encountered in existing systems or protocols and on the exploration of new approaches. This research framework at the interface of a specific network context and a challenging application domain, induces a close interaction, both with the underlying network level as well as the application level. Our methodology is based on a deep evaluation of the functionalities and performance of high speed infrastructures and on a study of the high end and original requirements before designing and analyzing new solutions. RESO gathers expertise in the design and implementation of advanced high performance cluster networks protocols, long distance networks and Internet protocols architecture, distributed systems and algorithms but also scheduling theory, optimization, queuing theory and statistical analysis. This background work provides the context model for innovative protocols and software design and evaluation. Moreover, the proposals are implemented and experimented on real or emulated local or wide area testbeds, with real conditions and large scale applications.

2.2. Context

Wavelengths multiplexing and wavelengths switching techniques on optical fibers allow core network infrastructures to rapidly improve their throughput, reliability and flexibility. Links of 40 gigabits per second are now available and 100 gigabits per second are emerging. New technologies like 10 Gigabit/s Ethernet or 10Gigabit/s Myrinet is also driving the increase of bandwidth in local area networks. These improvements have given the opportunity to create high performance distributed systems that aggregate storage and computation resources into a virtual and integrated computing environment. During a decade lot of researches and developments around the concept of grid and utility computing have underlined the strength of this approach. Today, the communication, computation and storage aspects of the Internet tend to converge. They are combined with the deployment of ultra high capacities interconnection networks with predictable performance and the emergence of coarse Grain Web Servers like Google, Yahoo, Amazon providing the content, control, storage and computing resources for the users. All these trends will strongly influence the development of the future Internet. They raise major research issues in networking and services, requiring a new vision of the network of networks and its protocol architecture. Indeed, the current Internet stack (TCP/IP) and its associated simple network management protocol is not consistent with the evolution of the network infrastructure components and its use by emerging services which aim to deliver supercomputing power available to the masses over the Internet. The coordination of networking, computing and storage requires the design, development and deployment of new resource management approaches to discover, reserve, co-allocate and reconfigure resources, schedule and control their usages. The network is not only a black box providing pipes between edge machines, but is becoming a vast cloud increasingly embedding the computational and storage resources to meet the requirement of emerging applications. These resources will be located at important crossroads and access points throughout the network. During the last few years, we have seen that the distinction between packet forwarding and application processing has become blurred. The network community now starts to worry not only about forwarding packets without regard to application semantics, but is increasingly trying to exploit new functionalities within the network to meet the requirements of application. Reciprocally, distributed systems and applications have traditionally been designed to run on top of the Internet, and to take the architecture of the Internet as given. Although the convergence of communication and computation at every level appears to be natural, it is still very difficult to efficiently explore the full range of possibilities it can bring. Most of

the proposals exploiting this convergence break the initial design philosophy of the Internet protocol stack (end-to-end argument for example), or if implemented in the application layer present lot of performance, resilience and scalability issues. Recently, ambitious research programs like FIND (NSF) or FIRE (EU) have been launched towards the design of a new protocol architecture for Future Internet and solving the critical issues facing the current architecture. We think that the Internet re-design raises the opportunity to better understand and assess higher-level system requirements, and use these as drivers of the lower layer architecture. In this process, mechanisms that are implemented today as part of applications, may conceivably migrate into the network itself, and this is one of main driver of the researches of RESO. One of the key challenge for large deployment of new high end applications in the Internet is the provisionning of a secure, flexible, transparent and high performance transport infrastructure for data access and processing. Consequently, future high-speed optical networks are addressed not only to support the accelerating and dynamic growth of data traffic but also the new emerging network requirements such as fast and flexible provisioning, QoS levels, and fast recovery procedures of such data intensive applications.

2.3. Research area

The use of networks for on-demand computing is now gaining in the large Internet, while the optical transport layer extends to the edge (fiber to the home). Enabling ultra high performance machine to machine communications lead then to new bandwidth and network resource sharing paradigms. RESO is investigating several issues such as quality of service, transport protocols, energy efficiency, traffic metrology , traffic modeling and network resource scheduling to deliver the emerging traffic in a timely, efficient, and reliable manner over long distance networks. In particular, RESO focuses of key issues such as :

- where and how to integrate the autonomy required to manage and control high speed networks at large scale?
- how dealing with the cost of resource and network virtualization on communication performance?
- which type of congestion control should be used in the context of a large scale deployment of high speed access networks (fiber to the home)?
- is the grid and high speed networks traffic self similar as the Internet traffic is? What are the critical scales? In which sense is self similarity harmful?
- how to efficiently share the bandwidth in a network that interleaves multimedia applications and computing applications?
- how to improve the interactions of message based communications or interactive traffic and transport layer in wide area networks?

To address some of these issues, our work follows four major research axes :

- Axis 1 : Optimized protocol implementations and networking equipements
- Axis 2 : Quality of Service and Transport layer for Future Networks
- Axis 3: High Speed Network's traffic metrology, analysis and modelling
- Axis 4: Network Services for high demanding applications

A large part of the axis 2 and axis 3 research topics is integrated in the ADR *Semantic Networking* of the common INRIA Bell Labs laboratory we are animating. The motivation of our research work in the common lab is to build and to exploit the knowledge that comes along with traffic. The goal is to act in a better way and to make better decisions at router and network level. The knowledge that comes part of traffic is what we refer to the "semantics" of traffic. The main topics we are exploring in this research axis of the common laboratory are

- Traffic identification and classification
- Traffic sampling
- Flow analysis
- Flow scheduling
- Sampling-based scheduling
- Flow-based routing

2.3.1. Axis 1 : Optimized protocol implementations and networking equipements

In this research axis we focus on the implementation and on the optimization of the mecanisms and process within networking devices. Since several years, virtualization of the operating system is used in end system to improve security, isolation, reliability and flexibility of the environments. These mechanisms become a must in large scale distributed system. In our research axis1 we explore how these mechanisms can be also adapted and used in data transport networks and specifically in switching and routing equipments.

However, virtualization introduces an overhead which must be integrated to system performance models in order to forecast their behavior. Lot of performance problems on end systems but also on router's data plane have to be studied and solved to make the virtualization approache viable. Investigating these issues is one of the goals in this research axis.

On an other hand, the key enabling factor of new network services is programmability at every level; that is the ability for new software capabilities to self-configure themselves over the network. We explore the concept, "dynamic programming enablers" for dynamic service driven configuration of communication resources.

In this research axis we also explore the integration of context-awareness functionality to address two important issues : reliability of communications and energy comsumption.

This direction is supported mainly by the EU FP7 "Autonomic Internet" project (2008-2010), with the INRIA "Action de Recherche Concertée" untitled "Green-NET" (2008-2010). The ANR HIPCAL (2007-2009) grant helps our studies around network virtualisation.

2.3.2. Axis 2 : Quality of Service and Transport layer for Future Networks

The goal of this axis is to guarantee quality of service in machine/user to machine/user communication while using efficiently the resources of the future networks. The two problems that are tackled here are: i) dynamic bandwidth sharing and congestion control in Future Internet and ii) control and flow management in semantic networks.

In this research axis, we focus on the three following questions:

- 1) which type of congestion control and transport protocol should be used in the context of large scale deployed high speed networks (fiber to the home, for example)?
- 2) how to efficiently share, but also dynamically provision the bandwidth of a network dedicated to computing tasks?
- 3) is the "flow-aware" approach a valuable solution to solve the end-to-end quality of service issue in the very high speed Future Internet?

2.3.3. Axis 3 : High Speed Network's traffic metrology and statistical analysis

Metrology of wide-area computer networks (i.e. the deployment of a series of tools allowing for collecting relevant information regarding the system status), is a discipline recently introduced in the context of networks, that undergoes constant developments. In a nutshell, this activity consists in measuring along time, the nature and the amount of exchanged information between the constituents of a system. It is then a matter of using the collected data to forecast the network load evolution, so as to anticipate congestion, and more widely, to guarantee a certain Quality of Service, optimizing resources usage and protocols design.
From a statistical signal processing viewpoint, collected traces correspond to (multivariate) time series principally characterized by non-properties: non-gaussianity, non-stationarity, non-linearities, absence of a characteristic time scale (scale invariance). Our research activity is undertaking the development of reliable signal analysis tools aimed at identifying these (non-)properties in the specific context of computer network traffic. In the course, we intend to clarify the importance of granularity of measurements.

Another challenge in network metrology is the effectiveness of packet sub-sampling. It means, to collect only a fraction of the overall traffic (supposedly redundant), and to study the possibility of inferring from that partial measurement, the most complete information about the system. Non trivial questions as, which fraction, which sub-sampling rule, adaptativity of this latter, smart sampling, statistical inference, open up a broad scope of investigation.
In this research axis, we focus on the two following questions:

- how does the traffic statistical properties really impact Quality of Service (QoS)?
- how to identify and to classify, in real time, transiting flows, according to a sensible typology?

Within the framework of the common laboratory between INRIA and Alcatel-Lucent, axis, "Semantic networking" brings in a new field of metrology research in RESO.

2.3.4. Axis 4: Network Services for high demanding applications

In strong interaction with the three fundamentals axes, this axis focuses on the application of the solutions to the grid context and on their implementation in a real environment such as the national research instrument Grid5000. Indeed, we believe that the precise structure of future applications and services is difficult to design without building large scale instruments and systems for real use based on real and high-performing hardware. Therefore, in this research axis we develop prototypes and deploy them within the Grid5000 testbed. For example, we design special measurement and routing systems at the edge of each Grid5000 site to explore new approaches or difficult problems alike. Topics that are investigated in this axis are strongly focusing the usage and the evolution of the Grid5000 instrument:

- Studies on the interactions of MPI and transport layer in wide area networks,
- Design, development and evaluation of a dynamic bandwidth provisionning service,
- Studies on the virtual private execution infrastructure concept for grid and cloud computing environments.
- Large scale deployment and evaluation of a high speed network measurement infrastructure.

RESO pursues researches for improving communications in grid environments. Thanks to systematic experiments of the behavior MPI in large scale environment, we merge optimizations of current implementations and propose new optimizations in the communication layers in order to execute more efficiently MPI applications on the Grid. We also study the impact of using TCP protocol for WAN communications (inter-site communications in the grid) and its interactions with MPI applications.

This research direction is mainly supported by FP6 EC-GIN grant, Grid'5000-ALADDIN initiative, CARRIOCAS project , HIPCAL project, JSPS-NEGST project. The DSLAB and OGF-Europe project provide resources.

2.4. Application domains

RESO applies its research to the domains of high end applications, distributed computing and to Grid and Cloud communications in particular.

Grid computing aims at bringing together large collection of geographically distributed resources (e.g., computing, storage, visualization, etc.) to build on demand very high performance computing environments for computing and data-intensive applications. These large scale cybernetic infrastructures gain increasing attention from a broad range of actors: from research communities to computer providers, large companies, and telecommunication operators (telcos). Whereas grids have been widely used in the scientific community, they are now moving into the commercial environment through the concept of Cloud computing solutions. Cloud computing fits a re-centralization scenario which offers suitable business and security model for large scale distributed resource sharing. Telcos are now moving toward infrastructure sharing and grid computing. Different scenarios for telcos are envisioned: telcos (1) deploy grids internally, e.g. for rapid dynamic service provisioning to new customers; (2) link different sites via VPNs; (3) act as a service broker. These scenarii are

explored with industrial partners. Researches conducted these last years reveal that grid technology raises new challenges in terms of network optimisation as well as of protocol architecture and of transport paradigms. We believe that a broad deployment of the grid and cloud technology can modify and influence the design of the future Internet as other emerging communicating applications. RESO design network services and network middleware, to simplify the programming and to optimize the execution of high end communicating applications while fully exploiting the capacities of the evolving networking infrastructure.

2.5. Methodology

The RESO approach relies on a methodology based on a three-steps cycle: 1) a fine analysis of limitations encountered in existing protocols (mainly TCP/IP), 2) the exploration of disruptive solutions, 3) the theoretical and experimental evaluation of these proposals. This research focuses an heavily ossified research object (the Internet protocol architecture) and lies between a challenging emerging application domain on a specific network context. These factors induce a close interaction with both the application level and the underlying network level as well as a deep technical and scientific knowledge of protocols and network equipments. The methodology is then based on a continuous study of the high end and original requirements and on experimental evaluation of the functionalities and performance of emerging dedicated high speed infrastructures. RESO gathers expertise in advanced high performance local and cluster area networks protocols, in distributed systems and algorithmics, in protocol and protocol architecture design, in long distance networking, in time series statistical analysis, in estimation theory and in performance evaluation. This background work provides the basis for innovative protocols and software design. Moreover, we implement and experiment our proposed prototypes on real, emulated local or wide area testbeds with real conditions and large scale applications.

2.6. Goals

RESO aims at providing software solutions but also original processes for high performance and flexible communications on very high speed networking infrastructures and for an efficient exploitation of these infrastructures. The goal of our research is to provide analysis of the limitations of the current communication and network software and protocols designed for standard networks and traditional usages, and to propose optimization and control mechanisms for the end-to-end performance, quality of service, energy efficiency and resource optimization. RESO explores original and innovative end-to-end transport services and protocols that meet the needs of high end applications. These solutions must scale in increasing bandwidths, heterogeneity and number of flows.

RESO studies high speed networks and their traffic characteristics, high end applications requirements, creates open source code, distributes it to the research community for evaluation and usage and help in shortening the wizard gap between network experts and novices. The long term goal is also to contribute to the evolution of protocols, standards and networking equipments, prompting the introduction of metrology as an intrinsic component of high-speed networks. An important effort is naturally dedicated to the dissemination of these new approaches.

2.7. Summary of the main contributions of the team in 2009

During this year, RESO team had main contributions in the following fields:

2.7.1. Axis 1 : Protocol implementations and networking equipments

- Exploration of the data plane virtualization cost and opportunities in software routers (VXRouter);
- Design of an Autonomic Network Programming Interface
- High availability for clustered stateful network equipments
- Models and software frameworks for energy efficiency in large scale distributed systems

2.7.2. Axis 2 : End-to-end Quality of Service and Transport layer

- Optimization algorithms for network resource sharing in very high speed networks (BDTS).
- Algorithms for dynamic bandwidth provisioning based on flows scheduling and aggregation
- Design of a language for specifying virtual infrastructures (VXDL)
- Study on network and system virtualisation for virtual private infrastructure creation and usage.
- Distributed algorithms for bandwidth sharing in mobile or very high speed environments.
- Analysis of opportunity of extending the size of transfer units (XLFrames).
- Analysis of flow scheduling and sampling-based scheduling
- Analysis of flow-based routing

2.7.3. Axis 3: High Speed Network's traffic metrology and statistical analysis

- Derivation of a maximum likelihood estimator of heavy tail distribution index from censored data. This estimator was applied to a set of incomplete flows reconstructed from a packet sub-sampled high speed network traffic, to get the tail index of the corresponding flow size distribution.
- Extension of Taqqu's relation relating heavy-tailed flow size distributions to long range dependence, to take into account the correlation between flow size and flow throughput.
- Refine the relations between observed statistical scaling properties of Internet traffic and performance measures (delay, loss rate) as QoS metrics.
- Identification on a long-lived TCP flow, of a new type of scale invariance property with multifractal support. Demonstration of a (almost sure) large deviation principle guaranteeing the identifiability of the corresponding multifractal spectrum from a finite size real trace. Adaptation to TCP markovian models to characterize fairness and sources' synchronization.

2.7.4. Axis 4: Network Services for high demanding applications

- Design of a network service for dynamic bandwidth provisioning in very high speed environments;
- Specification of a network resource scheduling, virtualization and reconfiguration component in a service oriented approach (Carriocas)
- Design and development of MPI5000, a communication layer for MPI over wide area network
- Traffic monitoring of LCG (CERN LHC Grid) 10Gb/s link at packet resolution to characterize Grid traffic;
- Design and development of the HIPerNet software for network-aware virtual cluster management tool.
- Pursue the collaborations for the development and usage of the GRID5000 international optical interconnections to Netherlands (DAS3) and Japan (Naregi) in collaboration with RENATER;
- Design and development of a metrology infrastructure for fine grain traffic monitoring in Grid5000.

3. Scientific Foundations

3.1. Optimized Protocol implementations and networking equipements

Participants: Jean-Patrick Gelas, Olivier Glück, Laurent Lefèvre, Pascale Vicat-Blanc Primet, Jean-Christophe Mignot, Ludovic Hablot, Sébastien Soudan, Fabienne Anhalt, Olivier Mornard.

The initial goal of the DARPA Internet Architecture was to develop an effective technique for multiplexed utilization of existing interconnected networks. Robustness was the first priority which strongly colored the design decisions within the Internet architecture. An architecture primarily for commercial deployment would have clearly placed the resource management at the beginning of the priority list. Some of the most significant problems with the Internet today relate to lack of sufficient tools for distributed management. For example, in the large Internet being currently operated, routing decisions need to be constrained by policies for resource usage. Today this can be done only in a very limited way, which requires manual setting of tables. This is error-prone and at the same time not sufficiently powerful. A key enabling factor of new services and protocols is then the ability for new software capabilities to self configure themselves over the network. Moreover, in the Future Internet, only trusting nodes should be able to communicate at will. Nodes should also be protected from nodes they do not want to communicate with. Virtualization, context-awareness and energy efficiency are then promising concepts for Future networks. However their potential and limits in the context of dynamic and self-organized high speed networks have to be studied.

Since several years, virtualization of the operating system is used in end system to improve security, isolation, reliability and flexibility of the environments. These mechanisms become a must in large scale distributed system. Virtualized resources is a new way of sharing in which group of users or activities (or trusting nodes) are given static shares, and only within these groups there is dynamic sharing. Virtual networks present ideal vantage point to monitor and control the underlying physical network and the applications running on the virtual trusting nodes. How virtualization can be also adapted and used in data transport networks and specifically in switching and routing equipments is an open question. For example, virtualization introduces an overhead which must be integrated to system performance models in order to forecast their behavior. Lot of performance problems on end systems but also on router's data plane have to be studied and solved to make the virtualization approach viable. Investigating these issues is one of the goals in this research axis.

On an other hand, the key enabling factor of new network services is programmability at every level; that is the ability for new software capabilities to self-configure themselves over the network. We explore the concept, "dynamic programming enablers" for dynamic service driven configuration of communication resources. Dynamic programming enablers apply to an executable service that is injected and activated into the network system elements to create the new functionality at runtime. The basic idea is to enable trusted parties (users, operators, and service providers) to activate management-specific service and network components into a specific platform. We study mechanisms and infrastructures required to support these components. We aim at providing new functionality to services using Internet facilities, addressing the self-management operations in differentiated and integrated services. The goal is the enhancement of the creation and the management (customization, delivery, execution and stop) of Internet services.

In this research axis we also explore the integration of context-awareness functionality to address two important issues : reliability of communications and energy consumption.

Session awareness : Most of the NGN services involve a session model based on multiple flows required for the signaling and for the data exchange, all along the session lifespan. New service-aware dependable systems are more than ever required. Challenges to these models include the client and server transparency, the low cost during failure free periods and the sustained performance during failures. Based on our previous work with FT R&D ("Procédés de gestion de sessions multi-flux", N. Ayari, D. Barbaron, L. Lefèvre, France Telecom R&D Patent, June 2007), we continue to explore and propose session aware distributed network solutions which support the reliability mandatory to operators services (VOIP).

Energy awareness : Large scale distributed systems (and more generally next generation Internet) are facing infrastructures and energy limitations (use, cooling etc.). In the context of monitored and controlled energy usage, we plan to explore the proposal of energy aware equipments and frameworks, which allow users and middleware to efficiently use large scale distributed architectures.

We are developing solutions to dynamically monitor energy usage, inject this information as a resource in distributed systems and adapt existing jobs (OAR) and network (BDTS) schedulers to autonomically benefit from energy information in their scheduling decisions. This research is linked with experimental evaluation on Grid'5000 platform and inside the ALADDIN initiative.

3.2. Quality of Service and Transport layer for Future Networks

Participants: Pascale Vicat-Blanc Primet, Laurent Lefèvre, Sébastien Soudan, Romaric Guillier, DInil Mon Divakaran, Guilherme Koslovski, Isabelle Guerin-Lassous, Rémi Vannier.

Congestion control is the most important and complex part of a transport protocol in a packet switched shared network. The congestion control algorithm is then a key component which has to be considered to alleviate the performance problems in the future networks environments. TCP has shown a great scalability in number of users, but not in link capacity and link diversity. For example, TCP performance can be very low and unstable in data-center applications and interactive communications within high speed long distance networks infrastructures, like lambda grids environments. The conservative behavior of TCP with respect to congestion in IP networks is at the heart of the current performance issues faced when the traffic load is highly dynamic. On the application side, one can observe that traditional applications were originally characterized by very basic communication requirements related to performance, reliability and order. The rapid deployment of new heterogeneous network technologies has pushed the development of an important number of new multimedia applications presenting complex requirements in terms of delay, bandwidth constraints and tolerance to losses. These applications need specific mechanisms to adapt to network congestion or changing medium conditions. To solve this problem, protocol enhancements and alternative congestion control mechanisms have been proposed for very high speed optical networks, wireless networks and for multimedia applications (see PFLDNET conference series). Most of them are now implemented in current operating systems, but these protocols are not equivalent, and not all of them are suitable for all environments and all applications, moreover they may not cohabit well. Since a couple of years, the evaluation and comparison of new transport protocols received an increasing amount of interest (see IRTF TMRG and ICCRG groups). However, TCP and other alternatives are complex protocols with many user-configurable parameters, and a range of different implementations. Several aspects can be studied, and various testing methods exist. The research community recognizes that it is important to deploy measurement methods so that the transport services and protocols can evolve guided by scientific principles. Researchers and developers need agreed-upon metrics, a common language for communicating results, so that alternative implementations can be compared quantitatively. Users of these variants need performance parameters that describe protocol capabilities so that they can develop and tune their applications. Protocol designers need examples of how users will exercise their service to improve the design.

As the Internet has evolved from a research project into a popular consumer technology, it may not be reasonable to assume that all end hosts would fairly cooperate. Indeed, concerns were raised that the recently started deployment of non-IETF-approved high-speed TCP variants could lead to an "arms race" that would eventually have a detrimental effect on the overall performance of the Internet. As another example, commercial Internet accelerators can provide better performance for a single user at the expense of other users. In the future, expecting billions of Internet devices to fairly cooperate to prevent network congestion is overly optimistic. New bandwidth sharing approach have to be investigated.

Flow scheduling [3], based on the in-advance knowledge of resource requirement of an application or online estimation of these requirements can be applied. Signaling or real time flow analysis and also scalability issues have to be explored. Distributed and lagrangian relaxation-based solution for *bandwidth sharing* is also an interesting approach in Future Internet. This approach addresses well the dynamic feature, due to node mobility or traffic variation. However, some problems remain open. First, the sharing models are often very complex to compute, while still being inaccurate. Second, some parameters of allocation algorithms based on lagrangian relaxation are difficult to set, and are often obtained by trial ad error; and hence not optimized. Finally, the proposed solutions are often tested on home-made simulators that are far from being realistic.

We believe some network resource control has to be associated with the end to end flow control approach to offer better quality of experience in Future networks. Network resource control is classified into three time-scales: data, control, and management. Each time-scale corresponds with a level of aggregation : ‘data’ deals with packets; ‘control’ deals with aggregates of packets, i.e. flows; and ‘management’ deals with aggregates of flows. All three time-scales must be addressed, since they all affect the service perceived by users, and the ease and efficiency with which the network can be operated. The current Internet protocols do not well

address the control time scale, and do not consider packet aggregates. TCP deals with resource control at data time-scales; while routing protocols, such as BGP and OSPF operate at time-scales of the order of minutes or hours (management time-scale).

We propose to explore packet aggregates and address control timescale in the context of Future Internet not only for performance, but also for manageability and security purposes.

On an other hand, the optical fiber communication will be the predominant mechanism for data transmission in core network and may be also at the access. To address the anticipated terabit demands, dynamically reconfigurable optical networks are envisioned. This vision will be realized with the deployment of configurable optical components, which are now becoming economically viable. To meet the terabit challenge, network designers will enhance core functionality by migrating to devices equipped with tunable transceivers, optical cross-connects and optical add/drop multiplexers. Optical Cross-Connects (OXCs) becomes more and more, cheap, simple and controllable. The control-plane, traditionally in the hand of telco migrates progressively to the customers. Studying the interactions of components required to accomplish the tasks of bandwidth reservation, path computation and network signaling is an other goal.

3.3. High Speed Network's traffic metrology and statistical analysis

Participants: Pascale Vicat-Blanc Primet, Paulo Gonçalves, Thomas Begin, Patrick Loiseau, Matthieu Imbert, Damien Ancelin, Olivier Grémillet.

Tools for measuring the end-to-end performance of a path between two hosts are very important for transport protocol and distributed application performance optimization. Bandwidth evaluation methods aim to provide a realistic view of the raw capacity but also of the dynamic behavior of the interconnection that may be very useful to evaluate the time for bulk data transfer. Existing methods differ according to the measurements strategies and the evaluated metric. These methods can be active or passive, intrusive or non-intrusive. Non-intrusive active approaches, based on packet train or on packet pair provide available bandwidth measurements and/or the total capacity measurements. None of the proposed tools, based on these methods, enable the evaluation of both metrics, while giving an overview of the link topology and characteristics.

That is the reason why a metrology activity including data processing, statistical inference, time series and stochastic processes analysis, deemed important to embed in the main research realm of RESO. Our goal is for these analyses to become in the near future a plain component not only in the study and in the development of infrastructures and computing networks, but also in real-time resources identification and management.

Grids specificities, such as the cooperating equipments number and heterogeneity, the number of independent processes, the treatments, bandwidth and stock capacities, turn indispensable to revisit the algorithms, as well as the control and operating mechanisms, in order to reach appropriate and optimal performances.

To validate a priori hypotheses that sustain already investigated approaches (e.g. overlay, virtualizing network resources, distributing network treatments, middleware programming), we resort to metrology and to the statistical analysis of the collected data. Indeed, we believe that automatic identification of static and dynamic properties of network resources is a prerequisite for developing adequate, adaptive and self-reconfigurable solutions.

We ground our approach on our large scale, fully controllable and configurable experimental facility (Grid5000+MetroFlux [37]) to validate, to better understand and to extend anterior results that were either heuristically observed or theoretically derived. Conversely, we perform realistic experiments, under prescribed and reproducible conditions, to get new insights into the statistical specificities of internet traffic, and to precisely identify the role of the network parameters [19].

Difficulty dwells in reliable classifiers and estimators of statistical properties, from non-stationary and possibly incomplete traces. We address these issues by proposing signal processing techniques well-tailored to network traffic measurements [36]. To go beyond the statistically description of the Internet traffic, we aim at investigating their effects on networking equipments such as routers and switches. An empirical study, implying the development of a " realistic " traffic generator, along with comprehensive sets of experimental

measurements should allow us to achieve this goal. RESO also aims at carrying out complementary analytical studies, based on the definition of theoretical models, so as to gain new insights in the performance measures resulting from the application of " Internet-like " traffic into classical queueing systems. To tackle these issues, RESO acquires, with the arrival of T. Begin in September 2009, new technical skills on modeling, queueing theory and performance evaluation.

Finally, the great investment that has been granted to Grid5000 (and to the interconnections Grid5000-Osaka) will profitably be used providing us with a high-performance, heterogeneous and quite novel experimental setup to confront the proposed theoretical models with real traffic measurements.

3.4. Network services for high demanding applications

Participants: Pascale Vicat-Blanc Primet, Olivier Glück, Laurent Lefèvre, Jean-Patrick Gelas, Paulo Gonçalves, Lucas Nussbaum, Patrick Loiseau, Olivier Mornard, Sébastien Soudan, Ludovic Hablot, Romaric Guillier, Manoj Dahal, Aurélien Cedeyn, Oana Goga, Armel Soro.

The purpose of Computational Grids was initially to aggregate a large collection of shared resources (computing, communication, storage, information) to build an efficient and very high performance computing environment for data-intensive or computing-intensive applications [82]. But generally, the underlying communication infrastructure of these large scale distributed environments is a complex interconnection of multi-IP domains with non controlled performance characteristics. Consequently *the Grid Network cloud* exhibits extreme heterogeneity in performance and reliability that considerably affect the global application performance.

The performance problem of the grid network cloud can be studied from different but complementary view points.

- Measuring and monitoring the end-to-end performance helps to characterize the links and the network behavior. Network cost functions and forecasts, based on such measurement information, allow the upper abstraction level to build optimization and adaptation algorithms.
- Optimally using network services provided by the network infrastructure for specific grid flows is of importance.
- Modeling, managing and controlling the grid network resource as a first class resource of the global environment: transfer scheduling, data movement balancing, bandwidth reservation and dynamic provisioning...
- Creating enhanced and programmable transport protocols adapted to heterogeneous data transfers within the grid may offer a scalable and flexible approach for performance control and optimization.

In a grid environment, two key points in the communication layers need to be taken in consideration in order to execute efficiently high performance applications: the heterogeneity of high-speed interconnects composing the grid and the Wide Area Network used to achieve inter-site communications. We explore new mechanisms to improve the application performance when it executes on the grid. We study, in particular, how a MPI application can benefit, during one execution, of several high-speed networks at the same time. In particular, it implies to find a way to communicate efficiently between these heterogeneous interconnections. We also explore how to keep good performance execution when long-distance communications are necessary because the application is launched on multiple sites of the grid.

An efficient MPI implementation for the grid is one of our research topic in this axis with the aim of improving communications in grid environments. The MPI standard is often used in parallel applications for communication needs. Most of them are designed for homogeneous clusters, but MPI implementations for grids have to take into account the heterogeneity of high-speed interconnects composing the grid and the Wide Area Network used to achieve inter-site communications, in order to maintain a high performance level. These two constraints are not considered together in existing MPI implementations, and raise the question of MPI efficiency in grids. Our goal is to significantly improve the performance execution of MPI applications on the grid.

Finally, the resource mutualisation and sharing paradigm proposed by the Grid remains a very promising and powerful concept that we apply to network resource sharing at many levels. To explore new approaches or difficult problems alike, we design and deploy special shared network resource at the edge of Grid5000 sites [2]. The goal is develop "proof of concept" experiments for exploring, among others, traffic awareness, the buffer sizing problem, buffer and filtering "in route" approaches, router virtualization, multipath routing, and router assisted transport protocols and communication libraries (MPI5000).

4. Application Domains

4.1. Panorama

RESO applies its research to the domains of high performance Cluster and Grid communications. Existing GRID applications did already identify potential networking bottlenecks, either caused by conceptual or implementation specific problems, or missing service capabilities. We participated to the elaboration of the first GGF document on this subject [88] [87], [89]. Loss probability, important and incompressible latencies, dynamic behavior of network paths question profoundly models and technic used in parallel and distributed computing [81]. The particular challenge arises from a heavily distributed infrastructure with an ambitious end-to-end service demand. Provisioning end-to-end services with known and knowable characteristics in a large scale networking infrastructure requires a consistent service in an environment that spans multiple administrative and technological domains. The first bottleneck is often located at the interface between the local area network (LAN) and the wide area network (WAN). RESO conducted several actions in the field of Grid High Performance Networking in the context of the OGF, the European or National projects. These activities have been done in close collaboration with other INRIA and CNRS French teams (Grand Large, Mescal, Graal) involved in the GRID5000 and the Grid Explorer projects and other European teams involved in pfldnet and Glif communities. RESO joined the CARRIOCAS project which studies and implements a very high bit rate (up to 40 Gb/s per wavelength) network interconnecting super computers, storage systems and high resolution visualization device to support data and computing intensive applications in industrial and scientific domains. Our activities cover networking intelligence for high performance distributed applications.

Finally, the evolution of the Internet usage pushing the convergence of communication and computation at every level confirms our initial vision : the network should not be seen only as a black box providing pipes between edge machines, but as a vast cloud increasingly embedding the computational and storage resources to meet the requirement of emerging applications [6]. These resources are generally located at important crossroads and access points throughout the network. During the last few years we have seen that the distinction between packet forwarding and application processing has become blurred. The network community now starts to worry not only about forwarding packets without regard to application semantics, but is increasingly trying to exploit new functionalities within the network to meet the requirement of the application. Reciprocally, distributed systems and applications have traditionally been designed to run on top of the Internet, and to take the architecture of the Internet as given. The convergence of communication and computation at every level appears to be natural. It is however important to explore the full range of possibilities it can bring. Most of the proposals exploiting this convergence break the initial design philosophy of the Internet protocol stack (end to end argument for example), or if implemented in the application layer present lot of performance, resilience and scalability issues. We think that the Internet re-design raises the opportunity to better understand and assess higher-level system requirements, and use these as drivers of the lower layer architecture. In this process, mechanisms that are implemented today as part of applications may conceivably migrate into the network itself, and this is one of main driver of the researches of RESO and of our strong involvement in the new INRIA-BellLabs "Semantic Networking" research axis.

- RESO is closely involved in the evolution of the Grid 5000 testbed, and responsible for the networking aspects. Grid5000 is a national initiative aiming at providing a huge experimental instrument to the grid software and computer science research community. RESO participate to the INRIA development action ALADDIN. Participating to the design, deployment and usage of

such high performance experimental Network and Grid testbed allow us to gather a strong deep experience and unique expertise in high speed network and protocols exploration and tuning.

- RESO pursue the construction of an international community around Grid networks through the european EC-GIN project as well as with the OGF networking community.
- Through the ANR IGTMD project and is collaborating with the LCG and real physicists. A dedicated link deployed between IN2P3 (one of the largest computing center in France) and the FermiLab laboratory in Chicago, enable us to perform transport protocol experiments as well as traffic capture.
- RESO is bringing its expertise in Grids and Grid Networking to the CARRIOCAS project of the “pôle Ile de France System@tic". This collaboration enable us to explore the limits and the advantages of our previous results in the context of a 40Gb/s dynamically provisionable network.
- Through the ANR IGTMD project and is collaborating with the LCG and real physicists. A dedicated link deployed between IN2P3 (one of the largest computing center in France) and the FermiLab laboratory in Chicago, enable us to perform transport protocol experiments as well as traffic capture.
- Through the ANR HIPCAL project and the newly started PetaFlow project, RESO is collaborating with biology and medical imaging applications.

5. Software

5.1. BDTS: Bulk Data Transfer Scheduling Service

Participants: Sébastien Soudan, Dinil Mon Divakaran, Pascale Vicat-Blanc Primet.

The coordination of resource allocation among end points in controlled networks may require a service to transfer large data sets within time intervals. Such transfers commonly start some time after its request, use any variable bandwidth, and must complete before a deadline. BDTS is a software for Bulk Data Transfer Scheduling which gives users, applications or middleware the possibility to specify transfer requests as transfer jobs, and ensure a transparent control of them within controlled networks. BDTS manipulates and produces profiles which are step functions that express variable bandwidth assignment over time. BDTS incorporates a scheduler to divide the time windows of overlapping transfer jobs into multiple intervals. BDTS implements a multi-interval scheduling algorithm which minimizes the congestion factor of the network. It assigns independent bandwidth values to each transfer job at each interval, producing a bandwidth profile for each transfer job.

During this year, BDTS software has been enhanced to better fit the network model where users can provision their own infrastructure. This has been done first by defining a model of interaction between the different actors: users, service providers and network operator with the associated scheduling and network provisioning problem. And secondly by defining a solution based on a linear program to solve it. This part is now integrated in BDTS and will be reused to provision Carriocas’s pilot network. BDTS has been demonstrated in SC’08. BDTS is distributed under LGPL license and downloadable at : http://www.ens-lyon.fr/LIP/RESO/Software

- BDTS: Bulk Data Transfer Scheduling Service
- Type: software
- Scientific problem addressed: transfer delay predictability, high congestion avoidance due to massive data transfers
- Functional description: dynamic network bandwidth allocation, bulk data transfers scheduling
- worldwide diffusion, derived and resused in SRV (Alcatel-Inria)
- Status: free software
- State: prototype,
- APP : JBDTS version 1 du 15/12/2007: IDDN.FR.001.220025.000.S.P.2008.000.10700
- Transfer to LinKTiss Start-Up

5.2. FLOC: Flow control

Participants: Pascale Vicat-Blanc Primet, Sébastien Soudan.

This software solves the problem of enforcing a rate allocation profile made by an external bandwidth scheduler in a packet network. FLOC is the daemon present on end machine responsible to enforce the multi-interval bandwidth allocation profile received from a scheduler to a socket identified by a token and registered by user applications. FLOC changes GNU/Linux kernel's `qdisc` configuration according to current date and profile so that senders can only send at a given time at the rate they are allowed to. FLOC is distributed under LGPL license and downloadable at : http://www.ens-lyon.fr/LIP/RESO/Software

- Type: software
- Scientific problem addressed: Explicit flow rate control
- Functional description: Limitation and triggering of flow rate.
- nationale diffusion, deployment in GRID5000,
- reference implementation in EC-GIN community,
- Status: free software
- State: prototype,
- APP: version 0.12 du 17 fÃ©vrier 2009: IDDN.FR.001.290009.000.S.P.2009.000.10200
- Transfer to LinKTiss Start-Up

5.3. NXE: Network eXperiment Engine

Participants: Pascale Vicat-Blanc Primet, Romaric Guillier.

NXE (for Network eXperiment Engine) is a tool developed to be able to execute any particular scenario over any given topology. A scenario is defined as a sequence of dates at which networking events such the start of a new bulk data transfer occurs. This software automate the selection, deployment , configuration and activation on distributed resources of pieces of software required to execute a large scale and reproducible networking experiment. This software has been demonstrated during the SuperComputing'2007 event on the INRIA booth and is adapted and used for deploying different types of networking experiment (controlled measurement, network device evaluation, virtualization overhead measurement, HIPerNet validation...) within RESO team. A graphical user interface has been developed to simplify the usage of the automation tool. NXE is distributed under LGPL license and downloadable at : http://www.ens-lyon.fr/LIP/RESO/Software

- Type: software
- Scientific problem addressed: Automation of large scale networking experiment
- Functional description: Definition, configuration, deployment, run and analysis of a large scale experiment for protocol evaluation.
- nationale diffusion, deployment in GRID5000,
- Status: free software
- State: prototype,
- APP: version 1.0 de novembre 2008: IDDN.FR.001.030005.000.S.P.2009.000.10800
- Transfer to LinKTiss Start-Up

5.4. HSTTS: High Speed Transport protocols Test Suite

Participants: Pascale Vicat-Blanc Primet, Romaric Guillier.

HSTTS (for High Speed Transport protocol Test Suite) is software implementing a fixed set of data transfer scenarios. It is designed to help users evaluate the performance they ought to be able to get out of their networking infrastructure when they transfer data by using different types of transport protocols and services. This software has been presented during the SuperComputing'2007 event on the INRIA booth. BDTS is distributed under GPL license and downloadable at : http://www.ens-lyon.fr/LIP/RESO/Software

5.5. VXcore

Participants: Pascale Vicat-Blanc Primet, Sebastien Soudan, Philippe Martinez, Guilherme Koslovski, Fabienne Anhalt, Romaric Guillier.

The VXcore software is a set of independant modules that can be integrated in any network or infrastructure virtualisation framework.

5.5.1. VXCalendar

- Type: software module
- Scientific problem addressed: scheduling of virtual resources and infrastructures
- Functional description: Resource temporal database manager
- Status: proprietary software
- State: prototype - associated with a patent
- APP: version 1.0 du 15 mars 2009 : IDDN.FR.001.290012.000.S.P.2009.000.10800
- Transfer to LinKTiss Start-Up

5.5.2. VXtopology

- Type: software module
- Scientific problem addressed: management and control of virtual resources and infrastructures
- Functional description: Resource spacial database manager
- Status: proprietary software
- State: prototype - associated with a patent
- APP: version 1.0 du 15 mars 2009 : IDDN.FR.001.290012.000.S.P.2009.000.10800
- Transfer to LinKTiss Start-Up

5.5.3. VXScheduler

- Type: software module
- Scientific problem addressed: scheduling of virtual resources and infrastructures
- Functional description: Adaptation of virtual infrastructure request and scheduling.
- Status: proprietary software
- State: prototype - associated with a patent
- version 1.0 du 15 mars 2009 : IDDN.FR.001.290010.000.S.P.2009.000.10800
- Transfer to LinKTiss Start-Up

5.5.4. VXDL parser

- Type: software module
- Scientific problem addressed: Virtual infrastructures specifications and processing
- Functional description: interpretation and XML traduction of virtual infrastructures specifications
- Status: proprietary software
- State: prototype - associated with a patent
- version 2.0 du 20 mars 2009: IDDN.FR.001.260009.000.S.P.2009.000.10800
- Transfer to LinKTiss Start-Up

5.6. SRVdemonstrator

Participants: Pascale Vicat-Blanc Primet, Sebastien Soudan, Philippe Martinez, Manoj Dahal.

This software is demonstrating the dynamic bandwidth allocation service (Bandwidth on Demand). It has been developed in close collaboration with ALCATEL-LUCENT Bell Labs and integrates the ReSO's VXcore module; SRV demonstrator is MTOSI (TMF standard) compatible.

- Name: Scheduling, Reconfiguration and Virtualisation Software.
- Type: software
- Scientific problem addressed: Bandwidth on Demand (MTOSI compliant)
- Functional description: Scheduling, Reconfiguration and Virtualisation of Network resources for intensive computing environment.
- deployment within ALCATEL- LUCENT testbed
- Status: proprietary software, CARRIOCAS contract
- State: prototype,
- APP: on going
- Transfer to LinKTiss Start-Up

5.7. PATHNIF

Participants: Pascale Vicat-Blanc Primet, Romaric Guillier.

PATHNIF is exploring the end to end path of a user connection, detects any bottlenecks and propose workaround to improve end to end performance.

- Type: software
- Scientific problem addressed: Automation of large scale and high speed network bottleneck detection
- Functional description: Systematically analysis and evaluate the capacity of potential bottlenecks of an end to end network path.
- Status: proprietary software - associated with a patent
- State: prototype,
- APP: version 1.0 de mars 2009 : IDDN.FR.001.260002.000.S.P.2009.000.10800.
- Transfer to LinKTiss Start-Up

The associated PATHNIF patent has been patented (NÂ° 09/05285, le 4 Novembre 2009
Procédé de détection de goulet d'étranglement d'un chemin réseau: PATHNIF
Romaric Guillier, P. Vicat-Blanc Primet

5.8. HIPerNet v0.5: Cloud solution

Participants: Pascale Vicat-Blanc Primet, Olivier Mornard, Jean-Patrick Gelas.

HIPerNet engine is software implementing discovery, selection, allocation, scheduling and management of virtual private execution infrastructures over the Internet. HIPerNET v0.5 is focusing on virtual end-resource deployement and configuration. This software has been presented during the SuperComputing'2008 event on the INRIA booth. HIPerNet is distributed under GPL license and downloadable at : http://www.ens-lyon.fr/LIP/RESO/Software

5.9. SNE (Stateful Network Equipment)

Participant: Laurent Lefèvre [contact].

Joint work with Pablo Neira Ayuso from University of Sevilla (spain).

SNE is a complete library for designing a stateful network equipment (contains Linux kernel patch + user space daemon). The aim of the SNE library is to support issues related to the implementation of high available network elements, with specially focus on Linux systems and firewalls. The SNE library (Stateful Network Equipment) is an add-on to current High Availability (HA) protocols. This library is based on the replication of the connection tracking table system for designing stateful network equipments. SNE is an open source project, available on the web (CECILL Licence) at http://perso.ens-lyon.fr/laurent.lefevre/software/SNE.

5.10. ShowWatts: Real time energy consumption grapher.

Participants: Laurent Lefèvre, Anne-Cécile Orgerie, Jean-Patrick Gelas [contact].

Simple software used to display real time measures of energy consumed by processing nodes in a grid architecture. This software proposes a graphical interface connected to a set of powermeter devices. Graphical interface can display measures coming through a long distance secured network tunnel.

5.11. WattM : Monitoring framework for energy consumption of data centers

Participants: Laurent Lefèvre [contact], Anne-Cécile Orgerie, Jean-Patrick Gelas.

- Functional description: Monitoring and exposing electrical usage of large scale number of resources.
- usage and deployment in GRID5000
- Status: Open Source software
- State: prototype

5.12. XCP-i (Interoperable eXplicit Control Protocol)

Participants: Anne-Cécile Orgerie, Laurent Lefèvre.

XCP (eXplicit Control Protocol) is a transport protocol that uses the assistance of specialized routers to very accurately determine the available bandwidth along the path from the source to the destination. We propose XCP-i [85] which is operable on an internetwork consisting of XCP routers and traditional IP routers without loosing the benefit of the XCP control laws

An ns-2 module simulating XCP-i has been developed and will be available on the web. Based on a Linux kernel, a software XCP-i router is currently under development.

5.13. MPI5000

Participants: Ludovic Hablot, Olivier Glück, Jean-Christophe Mignot, Pascale Vicat-Blanc Primet.

MPI5000 is a communication layer between the application (MPI for example) and the transport protocol (TCP) which improves communications of distributed applications over wide area network in grids. For instance, MPI5000 reduces the impact of retransmissions and the impact of congestion window in such a context. MPI5000 is automatically and transparently executed without modifying the application. The general principle is to introduce proxies at the interface between the local network and the long- distance network to differentiate communications. These proxies allows to put forward the split of TCP connections in order to avoid losses and retransmissions on the long-distance links. This mechanism also allows to keep the congestion window closer to available throughput on the long-distance network. This work is detailled and evaluated in [62], and shows which applications can benefit from these optimisations.

5.14. Metroflux

Participants: Pascale Vicat-Blanc Primet, Paulo Gonçalves, Patrick Loiseau, Matthieu Imbert, Oana Goga, Armel Soro.

Metroflux system aims at providing researchers and network operators with a very flexible and accurate packetlevel traffic analysis toolkit configured for 1 Gbps and 10 Gbps speed links. These projects helped in the setup of the Metroflux prototype based on the GtrcNet FPGA-based device technology, on huge data storage resources for collected data and on specific statistical analysis tools.

- Name: Metroflux
- Type: hardware/software
- Scientific problem addressed: flow and packet-level traffic capture and analysis
- Functional description: combine a FPGA based device for packet capture, header extraction and time stamp with a high capacity storage and computing server for statistic analysis of very high speed network's traffics.
- worldwide diffusion, deployment in Grid5000 network, used by ALCATEL-LUCENT Bell Labs,
- reference implementation in INRIA Bell-Labs common laboratory,
- Status: free software and access
- State: prototype,
- Depot APP: on going

6. New Results

6.1. Optimized protocol implementation and networking equipements

6.1.1. Evaluation and optimization of network performance in virtual end systems and routers

Keywords: *system virtualization*, *traffic control*, *virtual router*.

Participants: Fabienne Anhalt, Jean-Patrick Gelas, Pascale Vicat-Blanc Primet.

Virtualization techniques are applied to improve features like isolation, security, mobility and dynamic reconfiguration in distributed systems. To introduce these advantages into the network where they are highly required, an interesting approach is to virtualize the internet routers themselves. This technique could enable several virtual networks of different types, owners and protocols to coexist inside one physical network. Systematic analysis and experiments of the cost of network virtualization have been conducted. Optimisation of scheduler have been proposed.

The evaluation of Xen 3.1's network performance with TCP flows on end-hosts and routers show that the multiplexing level (Dom0) was the bottleneck . We have shown that the performance could be improved by manipulating the scheduler parameters. Giving more weight to dom0 improves throughput and fairness.

The evaluation of Xen 3.2's network performance with TCP and UDP flows on end-hosts and routers have shown that the performance improved significantly compared to previous 3.1 version of Xen. Better throughput was obtained and dom0's CPU overhead decreases. No more unfairness exists. A stated bottleneck is the forwarding of small sized packets.

We have proposed a model of a virtual software router we have implemented with XEN and we have evaluated its properties. We show that the performance is close to the performance of non virtualized software routers, but causes an important processing overhead and unfairness in the share of the ressources. We study the impact of the virtual machine scheduler parameters on the network performance and we show that the module which is responsible of forwarding the packets between the virtual machines and the physical interfaces is the critical point of network communications. We analysed virtualization from the the data plane perspective. We explore the resulting network performance in terms of throughput, packet loss and latency between virtual machines, and also the correspondig CPU cost. The virtual machines act as senders or receivers, or as software routers forwarding traffic between two interfaces in the context of Xen. Our results show that the impact of virtualization on network performance is getting smaller with the successive Xen versions, making this approach a promising solution for data plane virtualization.The router migration with Xen 3.2 has been explored. The migration process is slowed down by the forwarding of flows by the virtual router, especially with TCP flows. It can take several minutes instead of several seconds in case of inactivity.

Exploiting our results on virtual software routers we start to investigate virtual router design. We first worked on a survey to attempt to sketch the evolution of the modern switch architectures. The survey covers the literature over the period 1987-2008 on switch architectures. Starting with the simple crossbar switch, we explore various architectures such as Output queueing, Input queueing, Combined Input/Output queueing, buffered crosspoint etc., that have evolved during this period. We discuss the pros and cons of these designs, so as to shed light on the path of evolution of switch architecture, in particular in the context of equipment virtualisation. We are curently working on the design and on the evaluation of a virtual switch.

6.1.2. High availability for clustered network equipments

Keywords: *fault tolerance*, *high availability*, *scalability*.

Participants: Laurent Lefèvre, Pascale Vicat-Blanc Primet.

A key component for improving the scalability and the availability of network services is to deploy them within a cluster of servers. The main objective of this work is to design a network traffic load balancing architecture which meets fine grained scheduling while efficiently spreading the offered network traffic among the available cluster resources.

- **A scalable architecture for balancing the offered network traffic**

  While a lot of researches have been conducted in the field of job and network load balancing, less interest has been granted to the impact of the granularity of the used mechanism on the reliable execution of the upper layer services. In fact, the currently used flow level network load balancing frameworks fail to achieve session awareness while efficiently spreading the offered network load among the available resources, typically, when the offered network session involves multiple and heterogeneous flows. Representative services range from familiar services like HTTP and FTP, to some recent services like multimedia streaming using RTSP/RTP/RTCP and Voice over IP using SIP. Our work aims to provide an architecture to efficiently balance the offered network sessions among the available processing resources within a cluster of servers.

- **A highly available architecture for balancing the offered network traffic**

  High availability allows service architectures to meet growing demands and to ensure uninterrupted service. In our work, we are interested in providing the continuous execution of the offered network sessions in case of failure of the legitimate entry point to the cluster as well as in case of the failure of the processing server inside the cluster. We noticed that current fault tolerant frameworks need to support consistent transport and application level failover mechanisms, and that transport layer protocols do not provide high availability capabilities. Indeed, TCP does not distinguish between a packet loss due to congestion, or a packet loss due to a server overload or due to a server/link failure. Thus, it reacts the same way to packet losses and to delays, by retransmitting the same segment to the same remote end point of the connection. Moreover, TCP tolerates short periods of disconnection not longer than a few RTTs. It disconnects the communicating hosts once specific timers expire. On the other hand, transport protocols rely on an explicit association between a service and its physical location for the wired Internet. Thus, when a host fails, the end-to-end flow terminates.

  In order to address this limitation, we proposed an active replication based system which enhances the reliability of the already established TCP flows. The proposed scheme is client transparent and does not incur any overhead to the end-to-end communication during failsafe periods, and performs well during failures. Parts of this work are protected by the Intellectual Property National Institute (INPI) patent disclosure No. FR0653546

  [79], [78], [75], [77], [76]

6.1.3. High availability for stateful network equipments

Keywords: *fault tolerance*, *high availability*.

Participant: Laurent Lefèvre.

Joint work with Pablo Neira Ayuso from University of Sevilla (Spain).
In operational networks, the availability of some critical elements like gateways, firewalls and proxies must be guaranteed. Some important issues like the replication of these network elements, the reduce of unavailability time and the need of detecting failure of an element must be studied. We propose the SNE library (*Stateful Network Equipment*) which is an add-on to current High Availability (HA) protocols. This library is based on the replication of the connection tracking table system for designing stateful network equipments.

Proposing stateful network equipments on open source systems is a challenging task. We propose the basic blocks (SNE library) for building a stateful network equipment. This library can be combined with high-availability protocols (CARP, Linux HA...). We focus on Linux system in order to provide software solutions for designing high-available solutions for NAT, firewalls, proxies or gateways equipments...This library is based on components located in kernel and in user space of the network equipment. First micro-benchmark of communications mechanisms with Netlink sockets have shown the effectiveness of our approach

6.1.4. XCP-i: a new interoperable XCP version for high speed heterogeneous networks

Keywords: *TCP*, *XCP*, *XCP-i*, *available bandwidth*, *congestion control*, *virtual XCP-i router*.
Participant: Laurent Lefèvre.

XCP (eXplicit Control Protocol) is a transport protocol that uses the assistance of specialized routers to very accurately determine the available bandwidth along the path from the source to the destination. In this way, XCP efficiently controls the sender's congestion window size thus avoiding the traditional slow-start and congestion avoidance phase. However, XCP requires the collaboration of all the routers on the data path which is almost impossible to achieve in an incremental deployment scenario of XCP. It has been shown that XCP behaves badly, worse than TCP, in the presence of non-XCP routers thus limiting dramatically the benefit of having XCP running in some parts of the network. In this work, we address this problem and propose XCP-i which is operable on an internetwork consisting of XCP routers and traditional IP routers without loosing the benefit of the XCP control laws.

XCP-i basically executes the next four steps to discover and compute a new feedback that reflects the state of the network where non-XCP routers are placed:

1. Discover where the non-XCP routers are in the data path.
2. Discover the upstream and downstream XCP-i routers of the non-XCP routers.
3. Estimate the available bandwidth where the non-XCP routers are placed.
4. Create a virtual XCP-i router that computes a new feedback using the estimated available bandwidth before.

The simulation results on a number of topologies that reflect the various scenario of incremental deployment on the Internet show that although XCP-i performances depend on available bandwidth estimation accuracy, XCP-i still outperforms TCP on high-speed links [85].

6.1.5. Autonomic Service Deployment in Next Generation Networks

Keywords: *autonomic network*, *programmability*, *service deployment*.
Participants: Abderhaman Cheniour, Jean-Patrick Gelas, Laurent Lefèvre.

RESO is involved in the FP7 Autonomic Internet project by focusing on autonomic service deployment solutions for large scale overlays.

Programmability in network and services encompasses the study of decentralised enablers for dynamic (de)activation and reconfiguration of new/existing services, including management services and network components. The challenge in Autonomic Internet FP7 project (AutoI) is to enable trusted parties (users, operators, and service providers) to activate management-specific service and network components into a specific platform. Dynamic programming enablers will be created that are applied to executable service code, which can be injected/activated into the system's elements to create the new functionality at runtime. Network and service enablers for programmability can therefore realise the capabilities for flexible management support required in AutoI.

RESO has proposed the ANPI : Autonomic Network Programming Interface which will support the service enablers plane of the AUTOI architecture. This interface is currently under devlopment with the support of other AUTOI partners (Hitachi Europe, University College of London, UPC Barcelona, Univeristy of Passau).

6.1.6. Energy-efficiency in computing and networking for large-scale distributed systems

Keywords: *Energy-awareness*, *Energy-efficiency*, *Grid monitoring*.

Participants: Marcos Dias de Assuncao, Alejandro Fernandez, Jean-Patrick Gelas, Isabelle Guerin-Lassous, Laurent Lefèvre, Anne-Cécile Orgerie.

High performance computing aims to solve problems that require a lot of resources in terms of power and communication. While an extensive set of research project deals with the saving power problem of electronic devices powered by electric battery, few have interest in large scale distributed systems permanently plugged in the wall socket. The general common idea is indeed that, when they are not reserved, the grid resources should be always available, so that they should always remain fully powered on.

The large-scale distributed systems are sized to support reservation bursts. So they are not fully used all the time. Between the bursts, some resources remain free, so we can save energy during these gaps. This is our first approach taken in this work: to save energy by shutting down nodes when they are not used. We use the same approach for high performance data transport: the high-speed links are not always fully used and we can turn off the Ethernet cards and switch ports off to save energy.

Understanding the characteristic usage and workloads of the large-scale distributed systems is a crucial step towards the design of new energy-aware distributed system frameworks. Therefore we have studied the Grid5000 platform usage over long periods of time.

The analysis of these usage traces lead us to propose an energy-aware reservation infrastructure (EARI) which is able to shut down nodes when they are idle. This infrastructure proposes several energy efficient solutions for a reservation made by a user: several energy-efficient possibilities for his reservation.Thus the user is able to choose among these “green” solutions and this leads to an aggregation of the reservations. This infrastructure also includes a prediction algorithm to anticipate the next reservation in order to avoid shutting down nodes that we will need to be restarted quickly.

So, our infrastructure is based on three mechanisms:

- switching on and off the nodes;
- reservation aggregation with green policiesand
- predictions of the next reservations.

This model has been validated over the Grid5000 traces by using a replay mechanism. The results are really encouraging and show that our infrastructure could make huge energy savings. This on/off model is a first step in our research on energy efficiency in computing and networking for large-scale distributed systems.

We are working on improving the prediction models with Alejandro Fernandez from University of Seville, Spain.

6.2. Quality of service and Transport Protocols for Future Networks

6.2.1. Bulk Data Transfer Scheduling and Dynamic Bandwidth provisionning

Keywords: *bulk data transfers*, *dynamic bandwidth provisioning*, *flow scheduling*, *optical networks*.

Participants: Pascale Vicat-Blanc Primet, Sébastien Soudan, Martinez Philippe.

As the Internet has evolved from a research project into a popular consumer technology, it may not be reasonable to assume that all end hosts would fairly cooperate. In this context we are investigating new bandwidth sharing approaches.

Since several year we focus on different form of flow scheduling.

In this work we propose to manage explicitely the movements of massive data set between end point. We formulate the bulk data transfer scheduling problem and give an optimal solution to minimize the network congestion factor of a dedicated network or an isolated traffic class. The solution satisfying individual flows time and volume constrains can be found in polynomial time and expressed as a set of multi-interval bandwidth allocation profiles. To ensure a large scale deployment of this approach, we propose, for the data plane, a combination of a bandwidth profile enforcement mechanism with traditional transport protocols.

We pursue our exploration of the Bulk Data Transfer Scheduling Service (developed by INRIA and UIBK in the framework of the EU EC-GIN project (IST045256)), which operates at the control timescale. This service schedules and forwards packet aggregates for improving the predictability of massive data set transfer time. This service introduces and exploits the time dimension and a fine-grain user control plane. It implements the virtual network resource reservation paradigm. Exploring this type of service-oriented network resource management at a large scale in a heterogeneous environment will help in understanding the limits and alternatives of this approach. This help to have a better insight on fundamental issues, such as how does the control-plane interact with the data-plane, and, how do the abstraction layers, session, transport, network (corresponding to different timescales and aggregation levels) interact. This will clarify the limits of the current abstractions and protocols architecture and validate the potential of new abstractions. The Bulk Data Transfer Scheduling approach is based on the in-advance knowledge of resource requirement of an application or online estimation of these requirements can be applied. Signaling or real time flow analysis and also scalability issues are explored.

On an other hand, the optical fiber communication will be the predominant mechanism for data transmission in core network and may be also at the access. To address the anticipated terabit demands, dynamically reconfigurable optical networks are envisioned. This vision will be realized with the deployment of configurable optical components, which are now becoming economically viable. SInce 2008, RESO integrates this new perspective to understand how this optical component interact with electronic component and how to configure, control and tune them with end computers in the context of our associated team with AIST (Japan) and the G-Lambda Project and in collaboration with Alcatel-Lucent in the context of the CARRIOCAS project.

CARRIOCAS projects studies and implements a high bit rate optical network capable of accommodating the requirements of data-intensive, high-performance distributed applications, in terms of bandwidth, quality of service guarantees, dynamic and automated service provisioning. The investigations are carried out under the constraints of supporting the applications on converged network infrastructures hosting other types of traffic. Distributed storage of massive volumes of data as well as collaborative high resolution remote visualization are under experimentation on a testbed. We analyzed the requirements brought by the applications on the network, compares different network architectures, presents the management architectures along with some resource selection optimization algorithms, and developed a demonstrator of the SRV (scheduling Reconfiguration and Virtualisaiton) component.

The SRV entity handles the service requests (bandwidth on demand for example), aggregates them and trigger the provisioning of different types of resources accordingly. We proposed to adapt to envisioned heterogeneous needs by multiplexing rigid and flexible requests as well as coarse or fine demands. The goal is to optimize both resource provisioning and utility functions. Considering the options of advanced network bandwidth reservations and allocations, the optimization problem has been formulated. The impacts of the malleability factor have been studied by simulation to assess the gain [22]. Simulations show that the temporal parameters of requests (deadline and patience) are the dominant criteria and that a small malleability can improve performance a lot.

6.2.2. Flow-aware networks

Keywords: *QoS*, *cross-protect*, *flow analysis*, *flow scheduling*, *flow-aware*, *game theory*, *quality of service*, *sampling*.

Participants: Pascale Vicat-Blanc Primet, Dinil Mon Divakaran, Olivier Grémillet, Paulo Gonçalves, Pierre-Solen Guichard, Isabelle Guérin Lassous.

This work is conducted in the context of INRIA Bell Labs and in close collaboration with MAESTRO team (Eitan Altman). Flows crossing IP networks are not equally sensitive to loss or delay variations because they do not have the same utility functions and the same final usage. Since several years, research effort has been devoted to solve the problem of the heterogeneous performance needs of the IP traffic. A class of solutions considers that the IP layer should provide more sophisticated services than the simple best-effort service to meet the application's quality of service requirements. Quality of service has been studied in IP networks in the context of multimedia applications. Re-thinking the fundamental paradigm of packet switching network in high speed networks is on the table. The idea is to go from a packet-level approach to a flow-oriented strategy. To cope with the scalability issues, we work on disruptive algorithms within equipements, and on fully distributed (or localized) solutions. Problems that need to be explored concern flow identification and classification (see also next research direction), flow admission control, flow routing, flow scheduling, interaction with transport protocols and system stability.

- **Flow identification and classification (see also next research axis)** The problem of traffic identification and classification has received considerable attention from the research community. Our interest here is on how to build an efficient global knowledge plane that can be used for taking decisions on traffic identifications locally. Traffic can be classified at application level, trying to identify the specific application associated with the traffic. For better flexibility, the behaviour of traffic can be used to classification. In this way, the classification itself can be independent of any new application type.

  Besides looking at traffic at a course level, it is also useful in analysing traffic at a finer level. Interesting decisions can be taken based on flow characteristics. Various important flow characterstics are size, age and rate. Decisions can be based on any one of these characteristics, or multiples of the same. In this direction an exhaustive study of current literature dealing with flow classification, with respect to their size or to their underlying application, has been achieved. This bibliography survey led us to retain:

  – the " Sample & Hold " technique and the " muti-stage filters" [80], for early on-line differentiation between elephants and mice;

  – a supervised classifier (C4.5) based on a set of 248 packet related parameters to discriminate among 12 application classes [86].

  In both situations, we intensively tested the proposed classifiers, in order *(i)* to assess their performance in terms of misclassification and confusion rates, and *(ii)* to check possible on-line implementations [83].

  Since performing per-packet measurement for per-flow analysis is computationally challenging, there is growing interest in obtaining useful information on flow characteristics using sampling. Sampling reduces the processing required to obtain flow statistics.

- **Flow scheduling** One of the actions that can be taken based on flow characteristics is scheduling. Tremendous amount of work has ben done in the area of scheduling of jobs, and of late, many researchers have applied this in the context of networking, to schedule flows.

  We incorporate the idea of sampling to schedule flows so as to induce less processing overhead. We propose a simple and practical scheduling strategy, as well as analyse the mean response time of flows when the classification is accurate, and when the classification is performed based on sampled information.

  Scheduling flows research lead to the development of many queueing models, capitalizing on the heavy-tail property of flow size distribution. Theoretical studies have shown that 'size-based' schedulers improve the delay of small flows without almost no performance degradation to large flows. On the practical side, the issues in taking such schedulers to implementation have hardly been studied. We looked into practical aspects of making size-based scheduling feasible in future Internet. In this context, we propose a flow scheduler architecture comprising three modules - Size-based scheduling, Threshold-based sampling and Knockout buffer policy - for improving the performance

of flows in the Internet. Unlike earlier works, we analyze the performance using five different performance metrics, and through extensive simulations show the goodness of this architecture.

- **Admission Control and flow routing** One of the goal of using flow-aware networking is to achieve quality of service guarantees at flow level, which is the relevant granularity level for more and more users, applications and services like video and audio streaming or image guided surgery over long distances. By performing implicit differentiation between types of traffic and providing best quality of service for all admitted flows even in overload situations, Cross-protect is promising. In this work, we have partly evaluated this architecture, and then we have proposed a further evaluation and improvements concerning the implementation and failure tolerance, like adaptive routing for example.
- **System stability and flow-aware approach** Size-based scheduling is advocated to improve response times of small flows. While researchers continue to explore different ways of giving preferential treatment to small flows without causing starvation to other flows, little focus has been paid to the study of stability of systems that deploy size-based scheduling mechanisms. The question on stability arises from the fact that, users of such a system can exploit the scheduling mechanism to their advantage and split large flows into multiple small flows. Consequently, a large flow in the disguise of small flows, may get the advantage aimed for small flows. As the number of misbehaving users can grow to a large number, an operator would like to learn about the system stability before deploying size-based scheduling mechanism, to ensure that it won't lead to an unstable system. In this study, we analyse the criteria for the existence of equilibria and reveal the constraints that must be satisfied for the stability of equilibrium points. Our study exposes that, in a two-player game, where the operator strives for a stable system, and users of large flows behave to improve delay, size-based scheduling doesn't achieve the goal of improving response time of small flows.

6.2.3. Integrating very large packets in networks

Keywords: *jumbo frames*, *queueing delay analysis*.

Participants: Pascale Vicat-Blanc Primet, Dinil Mon Divakaran.

This work is conducted in the context of INRIA Bell Labs and in close collaboration with MAESTRO team (Eitan Altman). Looking into the future, this work (part of the INRIA Bell Labs research) addresses the need for larger packet size, called XLFrame (XLF), for an Internet which is soon to witness stupendous amounts of traffic that have to be processed and switched at amplifying line rates. Increasing the size of the basic transporting unit in the Internet has far-reaching incentives that otherwise appear hard to achieve. For a variety of reasons, we foresee a future Internet that has both packets (sand) and XLFs (rocks). As a first step, we analyse the effects of introducing XLFs in a network, and find the following: the amount of packet-header processing is greatly reduced, while the fair multiplexing of XLFs with standard packets can be achieved using a more careful queue management in routers.

We also look into how we can make improvements through incremental research. In this direction, studying the effect of having large packets (of size $>>$ current MTU) in the current network is important as well as useful. Such packets are called *XLFrames* (or XLFs in short). Some of the motivating reasons for having XLFs in a network are: (1) to reduce power consumption at equipments by reducing the processing required, (2) achieving maximum throughput with increasing line rates, and (3) reducing per-packet cost involved in protocol processing and interrupt handling at the end-hosts.

In this work, we find that, though XLFs greatly reduces per-packet cost, flows using XLFs throttle packet-switched flows. Besides, XLF-switched flows experience higher loss rates. A solution to the unfairness comes in the form of Deficit Round Robin (DRR) scheduling that can be deployed at an equipment. DRR combined with ECN is seen to reduce the loss rates considerably.

6.2.4. Network virtualisation

Keywords: *VXDL*, *optical networks*, *resource virtualization*, *virtual infrastructure*.

Participants: Pascale Vicat-Blanc Primet, Guilherme Koslovski, Sebastien Soudan, Fabienne Anhalt, Romaric Guillier, Philippe Martinez.

With the expansion and the convergence of computing and communication, the dynamic provisioning of customized processing and networking infrastructures as well as resource virtualization are appealing concepts and technologies. Therefore, new models and tools are needed to allow users to create, trust and exploit such on-demand virtual infrastructures within wide area distributed environments. These ideas are investigated with the INRIA Planete, Grand Large and CNRS I3S and IBCP in the context of the ANR HIPCAL project. RESO is implementing them in the HIPerNet framework enabling the creation and the management of customized confined execution environments in a large scale context. We also investigate them in the context of the CARRIOCAS project. We are currently industrializing and transferring the knowledge, the know-how, the software and the associated patents to the RESO spinoff which will be launched in 2010. This year we explored different issues:

- **Network virtualisation and Security** In this context we proposed to combine network and system virtualization with cryptographic identification and SPKI/HIP principles to help the user communities to build and share securely their own resource reservoirs. Based on the example of biomedical applications, we study the security model of the HIPerNet system and develops the key aspects of our distributed security approach. Then we examined how HIPerNet solutions fulfill the security requirements of applications through different scenarios [46].
- **Network virtualisation and Application mapping** Optimally designing customized virtual execution infrastructure and mapping them in a physical substrate remains a complex problem. We propose to exploit the expertize of both the application and workflow developers to ease this process while improving the end user satisfaction as well as the infrastructure usage. We study in particular how this knowledge can be captured and abstracted in the intermediate VXDL language, our language for specifying and describing virtual infrastructures. Based on the example of a specific biomedical application and workflow engine, we study the different optimisation strategies enabled by such an approach. Comparison of executions ran on different virtual infrastructures managed by our HIPerNet system show how the exploition of the application semantic can improve the overall process. All the experiments are enjoing the Grid'5000 testbed substrate [35].
- **Network virtualisation and Dynamic resource provisionning** To adjust the provisioning of the resources to end-user demand variations, new infrastructure capabilities have to be supported. These capabilities have to take into account the business requirements of telecom networks. In this work we proposes service framework to offer Internet service providers a dynamic access to extensible *virtual private execution infrastructures*, through on-demand and in-advance bandwidth and resource reservation services. This *virtual infrastructure* service concept is being studied in the CARRIOCAS project and implemented thanks to the SRV component [22].
- **A language for virtual resources and interconnection networks description** VXDL was developed to help users, applications or middleware in the virtual components specification, model and representation. Basically, this language enables the description of virtual infrastructures which are composed by I) virtual resources, II) virtual network topology and III) virtual time. Using these three sets of features it is possible represents a virtual infrastructure composition (describing resources individually and in groups) detailing the network topology desirable (through links configuration and virtual routers) informing the execution timeline of each set of resources and links. Each component (resource or group) can have different parameters, allowing the configuration of size, software, hardware, location and functionality. In addition, VXDL is able to interact with some specific configurations for virtual infrastructures, as the definition of the virtual machines numbers that can be allocated in a physical resources; basically location (anchor) of a resource; and virtual routers usage. VXDL is defined using both BNF notation and XML standard, allowing its utilization in frameworks (or middleware) for management virtual environments. In this context, different systems can use VXDL for exchange information about the virtual infrastructures. This year we continue to develop and validate this langage. We are also working on it within OGF NML WG.

- **Validation of HIPerNET virtual infrastructure manager**: We investigate the benefit obtained with HIPerNET for the reservation and isolation of experimental slices on the Grid5000 test environment. The slice design, the integration of network description in a session reservation as well as the automatic deployment of all control software pieces are key aspects that are beeing investigated.

6.2.5. A Performance Evaluation Framework for Fair Solutions in Wireless Multihop Networks

Keywords: *history-dependent utility functions*, *performance evaluation*, *quality of service*.

Participants: Rémi Vannier, Isabelle Guérin Lassous.

Fairness in multihop wireless networks has received considerable attention in the literature. Many schemes have been proposed, which attempt to compute the “optimal" bit rates of the transmitting mobile nodes so that a certain fairness criterion is met. As the related literature indicates, there is a trade-off between fairness and efficiency, since fairness schemes typically reduce the channel utilization. Also, it is questionable whether certain fairness schemes have a positive or negative impact on the QoS of certain user services. So far, there has been limited research on the impact of the varying short-term allocations of these protocols, due to their inherent features and also nodes mobility, on the user-perceived QoS (and social welfare) for services of long duration.

In this work, we introduce an assessment framework, based on history-dependent utility functions that can be used as a holistic performance evaluation tool of these fairness schemes. These functions quantify the satisfaction that the ad hoc users obtain from the way their long-lived service sessions are allocated bandwidth, due to the behavior of the MANETs fair schemes. This way we can unambiguously compare the performance of various fair solutions whose maximization goals are inherently different (max-min fairness, proportional fairness, etc.). Finally, we demonstrate the usefulness of this framework by applying it on different protocols. This framework could also be used in any kind of networks.

6.2.6. Auction-based Bandwidth Allocation Mechanisms for Wireless Future Internet

Keywords: *auction theory*, *bandwidth allocation*, *heterogeneous wireless networks*.

Participants: Manos Dramitinos, Isabelle Guérin Lassous.

An important aspect of the Future Internet is the efficient utilization of (wireless) network resources. In order for the - demanding in terms of QoS - Future Internet services to be provided, the current trend is evolving towards an “integrated" wireless network access model that enables users to enjoy mobility, seamless access and high quality of service in an all-IP network on an “Anytime, Anywhere" basis. The term “integrated" is used to denote that the Future Internet wireless “last mile" is expected to comprise multiple heterogeneous geographically coexisting wireless networks, each having different capacity and coverage radius. The efficient management of the wireless access network resources is crucial due to their scarcity that renders wireless access a potential bottleneck for the provision of high quality services.

In this work, we propose an auction mechanism for allocating the bandwidth of such a network so that efficiency is attained, i.e. social welfare is maximized. In particular, we propose an incentive-compatible, effi- cient auction-based mechanism of low computational complexity. We define a repeated game to address user utilities and incentives issues. Subsequently, we extend this mechanism so that it can also accommodate multicast sessions. We also analyze the computational complexity and message overhead of the proposed mechanism. We then show how user bids can be replaced from weights generated by the network and transform the auction to a cooperative mechanism capable of prioritizing certain classes of services and emulating DiffServ and time-of-day pricing schemes. The theoretical analysis is complemented by simulations that assess the proposed mechanisms properties and performance. We finally provide some concluding remarks and directions for future research.

6.2.7. Adaptive Mechanisms for Bandwidth Sharing in Multihop Wireless Networks

Keywords: *QoS and Best Effort flows*, *bandwidth sharing*.

Participant: Isabelle Guérin Lassous.

In this work, we have designed a new cross-layer protocol which guarantees bandwidth of QoS flows by adpating effectively and dynamically throughput of best effort transmissions when it is necessary. Our protocol relies on a estimation of the available bandwith differentiated according to the type of packets (QoS or best effort data packets) and a proportional fair bandwidth sharing between best effort flows. With these features, this solution increases the acceptance rate of QoS flows while ensuring an efficient use of the remaining bandwidth between best effort as a fair sharing.

6.2.8. Towards a User-Oriented Benchmark for Transport Protocols Comparison in very High Speed Networks

Keywords: *High Speed networks*, *High Speed transport*, *Performance evaluation*, *Protocol Benchmark*, *TCP*.

Participants: Pascale Vicat-Blanc Primet, Romaric Guillier, Ludovic Hablot.

Standard TCP faces performance limitations in very high speed wide area networks, mainly due to a long end-to-end feedback loop and a conservative behaviour with respect to congestion. Many TCP variants have been proposed to overcome these limitations. However, TCP is a complex protocol with many user-configurable parameters and a range of different implementations. It is then important to define measurement methods so that the transport services and protocols can evolve guided by scientific principles and can be compared quantitatively. Users of these variants need performance parameters that describe protocol capabilities so that they can develop and tune their applications. The goal of this work to make some steps towards a user-oriented test suite and a benchmark, called HSTTS, for high speed transport protocols comparison. We first identified useful metrics. We then isolated infrastructure parameters and traffic factors which influence the protocol behaviour. This enabled us to define classes of representative applications and scenarios capturing and synthesising comprehensive and useful properties. We finally evaluate this proposal on the Grid'5000 experimental environment, and present it to the IRTF TRMG working group.

6.3. High Speed Network's traffic metrology and statistical analysis

6.3.1. Impact of the Correlation between Flow Rates and Durations on the Large-Scale Properties of Aggregate Network Traffic

Keywords: *aggregated network traffic*, *correlations*, *heavy-tailed flow size distributions*, *long range dependence*.

Participants: Patrick Loiseau, Paulo Gonçalves, Pascale Vicat-Blanc Primet.

Since the discovery of long-range dependence in network traffic in 1993, many models have appeared to reproduce this property, based on heavy-tailed distributions of some flow-scale properties of the traffic. However, none of these models consider the correlation existing between flow rates and flow durations. In this work, we extend previously proposed models to include this correlation. Based on a planar Poisson process setting, which describes the flow-scale traffic structure, we analytically compute the auto-covariance function of the aggregate traffic's bandwidth and show that it exhibits long-range dependence with a different Hurst parameter. In uncorrelated case, the model that we propose is consistent with existing models, and predict the same Hurst parameter. We also prove that pseudo long-range dependence with a different index can arise from highly variable flow rates. The pertinence of our model choices is validated on real web traffic traces.

6.3.2. Maximum likelihood estimate of heavy-tail exponents from sampled data

Keywords: *flow size*, *heavy-tail distributions*, *maximum likelihood estimation*.

Participants: Patrick Loiseau, Paulo Gonçalves, Pascale Vicat-Blanc Primet.

This work, published in the proceedings of ACM Sigmetrics 2009 [36], is a joint collaboration with the MISTIS team project.

In the context of network traffic analysis, we address the problem of estimating the tail index of flow (or more generally of any group) size distribution from the observation of a sampled population of packets (individuals). We give an exhaustive bibliography of the existing methods and show the relations between them. The main contribution of this work is then to propose a new method to estimate the tail index from sampled data, based on the resolution of the maximum likelihood problem. To assess the performance of our method, we present a full performance evaluation based on numerical simulations, and also on a real traffic trace corresponding to internet traffic recently acquired.

6.3.3. A new model revealing unexplored scale invariant properties of TCP throughput

Keywords: *large deviation principle*, *long-lived TCP flow*, *markov model*, *multifractal analysis*.

Participants: Patrick Loiseau, Paulo Gonçalves, Pascale Vicat-Blanc Primet.

This is a joint work with J. Barral (Prof. Univ. Paris 13).
Classical scaling laws in network traffic are commonly accepted as valuable indicators of the system's state. However, they are usually related to the dynamic of the connections, rather than to the TCP control mechanism itself. In this work we identified new scale-invariance properties of the throughputs of long-lived TCP RENO connections, and we proposed an adequate model able to reproduce these scaling laws. Our model relies on simple Markov chains for which we can theoretically prove, and experimentally verify, that they inherently possess the sought properties. We derived the corresponding large deviation spectra, which reveal reliable and sensitive fingerprints of the performance and fairness of competing TCP flows. Under controlled experimental conditions, we then demonstrated the flexibility and the versatility of this original approach on real TCP traces. In particular, we showed that the specificities of experimental conditions, such as synchronization or cross-traffic nature, can easily be taken into account and embedded in the model. We also presented experimental evidence that different TCP variants also exhibit scale-invariant properties of the same kind.

6.3.4. Traffic classification techniques supporting semantic networks

Keywords: *elephants and mice*, *semantic networking*, *supervised classification*.

Participants: Olivier Grémillet, Paulo Gonçalves, Pascale Vicat-Blanc Primet.

This work is part of our activity within Common Lab between INRIA and Alcatel-Lucent Bell Labs; it has been carried out in close collaboration with A. Dupas (Alcatel-Lucent).
The Semantic Networking concept has been introduced to solve the QoS, scalability and complexity challenges for the Future of Internet. Based on traffic awareness and flow entity, it contributes to an adaptive management of the network. The first important features are the better knowledge of the transported traffic and the processing time of the classification compatible with real-time operation. In this work, we present interesting techniques of classification for semantic networks. The detection of the biggest flows is first studied with Sample and Hold and multi-stage filter methods with successful classification probability. We hence analyze the impact of flow parameters on the application identification performance and classify them according to their accuracy. We finally discuss a potential hardware implementation architecture to validate the concept of semantic networking.

6.3.5. Multifractal analysis for in partum fetal-ECG diagnosis

Keywords: *Multifractal analysis*, *detection*, *heart rate variability*, *supervised classification*, *wavelets*.

Participant: Paulo Gonçalves.

Albeit grounded on different physical origins, it is not rare that distinct real-world problems share common mechanisms and/or formulations. This similitude naturally fosters the development of unified frameworks which can then match a wide range of applications. Moreover, as statistical signal processing frequently stands at the junction of several scientific domains, it is not surprising that statistical studies go beyond the scope of the application areas they were initially addressing. That is why the RESO team was led to participate to inter-disciplinary collaborations that do not straightforwardly relate to the main themes of the project activities, but which can capitalize with them.

This is a joint work with the Sisyphe team of the ENS Lyon Physics Lab and with the obstetric group of *Hôpital Femme Mère Enfant* of *Hospices civils de Lyon*.
In partum fetal suffering surveillance is a key task to prevent fetal and neonatal mortality due to asphyxia. This is partially conducted by monitoring and analyzing fetal Electrocardiogram recorded during the delivery phase: A strong variability measured on the corresponding heart beat time series indicates a normal process. Though satisfactory in practice, the currently used analysis/decision criteria lead to a high number of false positives. Multifractal analysis can be envisaged as a new tool to revisit time series variability analysis. Applied to data collected at Hospices Civils de Lyon, France, wavelet Leader based multifractal analysis is shown here to achieve a significant discrimination between the True Negative, True positive and False Positive classes of patients. This hence open promising tracks to decrease the number of False Positives achieved.

6.4. Network services for high demanding applications

6.4.1. Design and development of an MPI gateway

Keywords: *Grid*, *Grid5000*, *MPI*, *heterogeneity*, *high-speed interconnects*, *relays*.

Participants: Ludovic Hablot, Olivier Glück, Jean-Christophe Mignot, Pascale Vicat-Blanc Primet.

The MPI standard is often used in parallel applications for communication needs. Most of them are designed for homogeneous clusters but MPI implementations for grids have to take into account heterogeneity and long distance network links in order to maintain a high performance level. These two constraints are not considered together in existing MPI implementations and raise the question of MPI efficiency in grids. Our goal is to significatively improve the performance execution of MPI applications on the grid.

We have done a state of the art, a performance evaluation, understanding and tuning of four recent MPI implementations for the Grid : MPICH-Madeleine, GridMPI, OpenMPI and MPICH2. The comparison is based on the executions of pingpong, NAS Parallel Benchmarks and a real application of geophysics. These experiments take place on the national GRID'5000 testbed. We show that a tuning of both TCP protocol and MPI implementation are necessary to obtain good performances on the grid. We study the impact on application time execution of a long-way latency between two groups of 8 MPI tasks for each NAS parallel benchmark. Our experiments and tunings presented in [84] lead to the conclusion that GridMPI performs better results than the others and that executing MPI applications on a grid can be beneficial if some specific parameters are well tuned.

Based on these results, we propose a new transparent layer called MPI5000 and placed between MPI and TCP allowing application composed of several tasks to be correctly distributed on available node regarding the grid topology and the application scheme. Thus, our layer needs two data files: a file describing the grid topology including available nodes, both latency and bandwidth between the nodes and between sites; another file describing the application communication patterns with the size and the amount of messages sent between MPI processes. Using these two data files, our layer should realise an efficient placement of tasks on grid nodes.

Our layer also proposes to transparently slipt TCP connections between MPI processes in order to take into account the grid topology. This new architecture is based on a system of relays placed at the LAN/WAN interface. We replace each end-to-end TCP connection by three connections (two on the LAN between a node and a relay, one on the WAN between two relays). Thus, it allows a faster lost recovery on LAN as well as a reduction of memory used because the size of TCP buffers depends on RTT latency of the connection. Thanks to our architecture, we have proposed to use different TCP implementations for local and distant communications. The relays could also implement a different scheduling strategy of MPI messages : for instance, we could give priority to small messages (usually MPI control messages). Finally, as MPI applications are mostly using small messages, they are more penalised if the network is congestionned by large flows. Thanks to the communication aggreagtion between relays, we have showed that our acrhitecture allows to keep the congestion window closer to available throughput on the long-distance network.

This work is detailled and evaluated in [62], and shows which applications can benefit from these optimisations. We analyse for many points, the overhead and the benefits of the use of proxies. The theoritical analysis is supported by experiments. We conclude that for MPI applications that are using collective operations, the benefit on losses and retransmissions generally do not hide the overhead added by the splitting of the connections. Other applications benefit from this mechanism if they communicate sufficiently.

The implementation of MPI5000 is based both on a library between MPI and the operating system and on relays. Thus, the proposed architecture is independant of MPI implementations and is totally transparent for applications.

6.4.2. Development of a metrology platform on Grid5000

Keywords: *Gtrc-Net1*, *header extraction*, *metrology*, *monitoring*, *packet capture*.
Participants: Patrick Loiseau, Damien Ancelin, Aurélien Cedeyn, Matthieu Imbert, Romaric Guillier, Paulo Gonçalves, Pascale Vicat-Blanc Primet.

This activity is partially supported by the program GridNets-FJ (*Équipe associée*) between INRIA and AIST (Japan).
Researches in network traffic analysis embrace a large diversity of goals and are based on a variety of methodologies and tools. To have a better insight on the real nature and on the evolution of network traffic we argue that fine-grain analysis of real traffic traces have to complement simulations studies as well as coarse grain measurement performed by classical flow measurement systems. In particular, packet level measurements and analysis are needed. However, such methodologies are resource consuming and require very high performance devices to be operational in real high speed networks. In we present the *Metroflux* system which aims at providing researchers and network operators with a very flexible and accurate packet-level traffic analysis toolkit configured for 1 Gbps and 10 Gbps speed links. This system is based on the GtrcNet FPGA-based device technology and on specific statistical analysis tools. We show the potential and the facilities offered by the *Metroflux* system coupled with the *Grid5000* large scale experimental platform and the Network eXperiment Engine (*NXE*) we have developed. In we illustrate the application of *Metroflux* with the practical validation of the theoretical prediction relating self-similarity and heavy tails given by Taqqu theorem. We also illustrate several usages of this toolset, such as the investigation of conditions under which several traffic theories apply, as well as studies on traffic, protocols and systems interactions.

7. Contracts and Grants with Industry

7.1. INRIA actions

7.1.1. GRID5000: ADT Aladdin

Participants: Olivier Glück, Sébastien Soudan, Romaric Guillier, Ludovic Hablot, Laurent Lefèvre, Pascale Vicat-Blanc Primet, Paulo Gonçalves, Patrick Loiseau, Jean-Christophe Mignot, Aurélien Cedeyn.

ENS Lyon is involved in the GRID'5000 project, which is an experimental Grid platform gathering nine sites geographically distributed in France. ENS Lyon hardware contribution is done for now by two distinct set of computers. The Grid5000 of Lyon comprises now around 300 processors interconnected with a network of 500Mb/s Ethernet bisection and a 2Gb/s Myrinet interconnection for 64 nodes.

RESO is strongly involved in the choices of Grid5000's network components and architecture. Pascale Vicat-Blanc Primet is member of the national committee (comité de pilotage) of GRID'5000, of the Aladdin scientific committee, co-responsible of the Lyon site with Frederic Desprez, and coordinates networks aspects with Renater and RMU, Lyon's metropolitan network. Lyon site is nationaly recognized to gather the "networking expertize" with skilled researchers and engineers and dedicated networking equipments Metroflux, GNET10...). Working for the interconnection of the Grid5000 project at the international level, we are hosting the japanese Naregi project remote hosts and are accessing to dediated equipments within the Naregi testbed. We also participate to the ALADDIN ADT. Oana Gona, funded by the Aladdin ADT, is designing and developing an open traffic measurement and analysis infrastructure for the Grid5000 testbed. Aurélien Ceyden is member of the national technical committee of GRID'5000. Year funding: 60 K euros

7.1.2. INRIA ARC GREEN-NET

Participants: Laurent Lefèvre, Jean-Patrick Gelas, Anne-Cécile Orgerie.

The GREEN-NET is a Cooperative Research Action (ARC : Action de Recherche Cooperative) supported by INRIA. This project explores the design of energy-aware software frameworks dedicated to large scale distributed systems. These frameworks will collect energy usage information and provide them to resources managers and schedulers. Large scale experimental validations on Grid5000 and DSLLAB platforms will be proposed. Laurent Lefèvre is leading the INRIA ARC GREEN-NET on "Power aware software frameworks for high performance data transport and computing in large scale distributed systems" which involved 4 partners : INRIA RESO, INRIA MESCAL (Grenoble), IRIT (Toulouse), Virginia Tech (USA). Thanks to the ARC GREEN-NET, Marcos Dias de Assuncao has begun a postdoc position in RESO team in order to work on the design and experimental validations of energy aware large scale distributed systems. Official ARC GREEN-NET webpage : http://www.ens-lyon.fr/LIP/RESO/Projects/GREEN-NET

7.2. INRIA Bell Labs common laboratory: Semantic Networking

Participants: Pascale Vicat-Blanc Primet, Isabelle Guerin-Lassous, Paulo Gonçalves, Thomas Begin, Olivier Grémillet, Dinil Mon Divakaran, Pierre-Solen Guichard, Marina Sokol.

During this year we conducted the following researches:

- State-of-the-art on the different aspects covered within the Semantic Networking Aspects. Particular focus on the X-protect approach of France Telecom R&D.
- Study of the impact of large and small flows in current networks and analyse on how to handle both in Semantic Networks
- Proposals of new ideas through INRIA/Alcatel-Lucent discussions that will lead to patents, in elephant flow monitoring and scheduling/control.
- Global Semantic Networking architecture and high-level view of Semantic node defined.
- Development of the 10Gb/s packet capture system. Trace of 10Gb/s traffic on a real production network captured. The fine-grain analysis of these data is ongoing.

Year funding: 120Keuros

7.3. CARRIOCAS

Participants: Pascale Vicat-Blanc Primet, Manoj Dahal, Romaric Guillier, Guilherme Koslovski.

Carriocas project studies and implements an ultra high bit rate (up to 40 Gbps per wavelength) network interconnecting super computers, storage servers and high resolution visualization devices to support data and computing intensive applications in industrial and scientific domains. The R&D activities cover high bit rate transmission systems, advanced networking intelligence, and high performance distributed applications. CARRIOCAS is a three year project started in October 2006 which aims to be an experimental step of the transition from local to external storage and computing systems. This transition is valuable to share the cost of powerful systems among several users, to provide scalable and resilient architecture through distributed resource and to enable virual collaborative working environments between different actors working on a same project. The following points are especially investigated:

- Supporting the high bandwidth requirements through the migration of networks from 10 gbp/s to 40 Gbps/s per wavelength in a cost effective way.
- Building architectural, protocol and algorithmic solutions able to provide to the network the agility to dynamically adapt to the application needs with a high level of automation and optimizsation, while taking into account the administrative and business constraints.
- Developing and demonstrating on a network testbed distributed applications bringing perfomance enhancements for concrete scientific and industrial needs.
- Investigating the definition and the associated business models of high added value services integrating computing, visualization, storage and network resources.

In this project, RESO in charge of the design and protoyping of the "Resource Scheduling Reconfiguration and Virtualisation - SRV" component. Year funding : 100Keuros

8. Other Grants and Activities

8.1. National actions

8.1.1. ANR HIPCAL

Participants: Pascale Vicat-Blanc Primet, Jean-Patrick Gelas, Olivier Mornard, Fabienne Anhalt, Guilherme Koslovski, Philippe Martinez, Lucas Nussbaum.

HIPerCAL studies a new paradigm (grid substrate) based on confined virtual private execution infrastructure for resource control in grids. In particular, we propose to study and implement new approaches for bandwidth sharing and end to end network quality of service guarantees. The global infrastructure (computers, disks, networks) in partitioned in virtual infrastructures (aggregation of virtual machines coupled with virtual channels) dynamically composed. These virtual infrastructures are multiplexed in time and space, isolated and protected. The goal of this project is to explore an approach in a break with current services-oriented principles developed in grids to jointly enhance the application portability, the communications performance control and their security. The project aims at providing a grid substrate based on end to end bandwidth reservation, control overlay, network and system virtualization, cryptographic identification principles. The proposal is to be validated and evaluated at different scales on the Grid5000 testbed with biomedical applications, demanding in security , performance and reliability. 10 to 1000 processors, links with 100Mb/s to 10Gb/s, few microseconds to 100ms will be involved in these experimentations. We aim at demonstrating the functional transparency, enhanced predictability and efficiency for applications offered by the HIPerNet approach. RESO has developed, deployed and tested the first version of the HIPerNet software on the Grid5000 testbed.

Year funding: 100Keuros

8.1.2. ANR PETAFLOW

Participants: Paulo Gonçalves, Pascale Vicat-Blanc Primet, Matthieu Imbert.

This ANR (Appel Blanc International) started in october 2009 and will end in september 2012. It is a collaborative project between the GIPSA Lab (Grenoble), MOAIS (INRIA Grenoble), RESO (INRIA Grenoble), the University of Osaka (the Cybermedia Center and the Department of Information Networking) and the University of Kyoto (Visualization Laboratory).
It is no falsehood to state that “current society and science attempt to deal with increasing amounts of data". Today, peta-scale data are commonly gathered as well as generated thanks to the continuous development of measurement technologies and computational resources in diverse fields of science and society. Efficient processing or generation of peta-scale data requires high performance computational (HPC) resources which should be made remotely accessible through long-distance high performance networking and might be represented thanks to interactive scientific visualization. Consequently, generation or processing of peta-scale data benefits from the emergence of adequate “Information and communication technologies (ICT)" with respect to high performance “computing-networking-visualization" and their mutual “awareness". In the current proposal, it is aimed to develop and validate such ICT solutions using a transnational high-speed research network between Japan and France connecting GRID5000 (France) to the Naregi (Japan) testbed. Data-transfer protocols are aimed to be validated on data obtained for a real scientific problem involving peta-scale data.

Due to the medical relevance as well as basic scientific interest, peta-scale data are obtained from HPC Computational Fluid Dynamics (CFD) simulations on a vector supercomputer (NEC SCX9 Japan) aiming to predict the airflow through the upper airways. In addition, CFD simulation outcome is used as an input for aero-acoustic computations (CAA) for prediction of noise production. High performance computing is needed in particular to predict fricative noise due to the requested accuracy of the flow field (up to 16kHz). The outcome of CFD and CAA simulations will be validated on flow and noise measurements on a suitable experimental setup (France). Besides the international transfer of the generated peta-scale data, scientific visualization of peta-scale data is aimed on a single PC as well as on a tiled display wall for 3D interactive reconstruction of the flow and noise data.

In summary, the proposed project aims to contribute to the state-of-the-art of HPC, networking, scientific visualization and their mutual interactions for peta-scale data, while at the same time it is aimed to contribute to basic research in the fields of CFD and CAA applied to flow through the upper airways. The current proposal can only be realized thanks to the joint efforts and resources of the French and Japanese partners involved which gathers specialists in networking, middleware, scientific visualization, HPC and upper airway flow modeling and noise production.

8.1.3. ANR DMASC

Participants: Paulo Gonçalves, Patrick Loiseau.

Started in october 2008, this ANR project, leaded by J. Barral (Sisyphe, INRIA Roquencourt), is a partnership between INRIA (Sisyphe and Reso), university Paris 12 and university Paris Sud (équipe d'accueil EA 4046 Service de Réanimation Médicale CHU de Bicêtre).
Numerical studies using ideas from statistical physics, large deviations theory and functions analysis have exhibited striking scaling invariance properties for human long-term R-R interval signals. These signals are extracted from electrocardiograms and represent the time intervals between two consecutive heartbeats. The scaling invariance measured on these empirical data are reminiscent of geometric fractal properties verified theoretically by certain mathematical objects (measures or functions), which are called (self-similar) multifractals. These numerical studies also reveal that the scaling invariance may have different forms, according to the fact that the patients have a good health or suffer from certain cardiac diseases. These observations suggest that a good understanding of multifractal properties of cardiac signals might lead to new pertinent tools for diagnosis and surveillance. However, until now, neither satisfactory physiological origin has been associated with these properties nor mathematical objects have been proposed as good models for these signals. It is fundamental for possible medical applications in the future to go beyond the previously mentioned works and achieve a deepened study of the scaling invariance structure of cardiac signals. This requires new robust algorithms for the multifractal signals processing; specifically, it seems relevant to complete the usual statistical approach with a geometric study of the scaling invariance. In addition, it is necessary to apply these tools to a number of data arising from distinct pathologies, in order to start a classification of the different features of the observed scaling invariance, and to relate them to physiological concepts. This should contribute to develop an accurate new flexible multifractal mathematical model whose parameters could be adjusted according to the observed pathology. It is also important to strengthen the information by performing the multifractal analysis of another fundamental signal in cardiology, namely the blood pressure, as well as the simultaneous multifractal analysis/modeling of the couple (R-R,Blood Pressure). This project aims at achieving such a program. It also proposes to contribute to explain the origin of the scaling invariance properties by developing a reduced order dynamical system, which shall describe the heart's electromechanical activity and simultaneously shall generate multifractal outputs in accordance with the R-R signals models. A 1-D model of cardiac fiber would be already very satisfactory. This aspect of the project is closely related to the delicate issue of understanding the link between multifractal phenomena and PDEs, another topic that will be investigated. The project team consists in six members representing four partners: two specialists of multifractal analysis, one specialist of cardio-vascular system modeling and PDEs control, one specialist of statistical signal processing and two physiologists (among which one cardiologist) specialists of cardio-vascular signals processing. The project will benefit of a wide data's bank of long term (24h) R-R interval signals already recorded in various clinical settings including diabetes, acromegaly and sleep apnea, and

a prospective data bank will be established in the field of medical intensive care unit, namely in patients presenting cardiovascular pathologies like heart failure, arterial hypertension and chock states. The data bank will include both R-R interval signals and arterial blood pressure signals.

Year funding: 2,5Keuros

8.2. European actions

8.2.1. AUTONOMIC INTERNET - 2008-2010

Participants: Laurent Lefèvre, Jean-Patrick Gelas, Abderhaman Cheniour.

Autonomic Internet (AutoI - FP7.ICT.2007.Call1-216404) project suggests a transition from a service agnostic Internet to service-aware network, managing resources by applying autonomic principles. In order to achieve the objective of service-aware resources and to overcome the ossification of the current Internet AutoI will develop a self-managing virtual resource overlay that can span across heterogeneous networks that can support service mobility, security, quality of service and reliability. In this overlay network, multiple virtual networks co-exist on top of a shared substrate with uniform control. The overlay will be self-managed based on the system's business goals, which drive the service specifications, the subsequent changes in these goals (service context) and changes in the resource environment (resource context). This will be realised by the successful co-operation of the following activities: autonomic control principles, resource virtualisation, enhanced control algorithms, information modelling, policy based management and programmability. RESO is mainly involved in the programmability of the AUTOI overlay by proposing an Autonomic Network Programming Interface which will support large scale service deployment. Laurent Lefèvre is leading the workpackage 5 on "Service Deployment". Official webpage : http://www.ens-lyon.fr/LIP/RESO/Projects/Autonomic_Internet/demo.html

8.2.2. OGF-EUROPE - 2008-2010

Participants: Laurent Lefèvre, Augustin Ragon, Pascale Vicat-Blanc Primet.

RESO participate in the OGF-Europe to reinforce the french participation to OGF standardization activities. We mainly concentrate our contribution on Telco interaction and Energy-efficiency in Grid context.

8.2.3. COST Action IC0804 on Energy efficiency in large scale distributed systems - 2009-2013

Participants: Laurent Lefèvre, Jean-Patrick Gelas, Anne-Cécile Orgerie.

The main objective of the Action is to foster original research initiatives addressing energy awareness/saving and to increase the overall impact of European research in the field of energy efficiency in distributed systems. The goal of the Action is to give coherence to the European research agenda in the field, by promoting coordination and encouraging discussions among the individual research groups, sharing of operational know-how (lessons-learned, problems found during practical energy measurements and estimates, ideas for real-world exploitation of energy aware techniques, etc.).The Action objectives can be summarized on scientific and societal points of view: sharing and merging existing practices will lead the Action to propose and disseminate innovative approaches, techniques and algorithms for saving energy while enforcing given Quality of Service (QoS) requirements. Laurent Lefèvre is Management Committee member and French representative in this COST action.

8.2.4. AEOLUS

Participants: Manos Dramitinos, Isabelle Guérin-Lassous, Rémi Vanier.

AEOLUS (Algorithmic Principles for Building Efficient Overlay Computers) is an IP project that has been started since September, 1st, 2005. The university of Patras (Greece) is the prime contractor. The goal of this project is to investigate the principles and develop the algorithmic methods for building an overlay computer that enables an efficient and transparent access to the resources of an Internet-based global computer. In particular, the main objectives of this project are:

- To identify and study the important fundamental problems and investigate the corresponding algorithmic principles related to overlay computers running on global computers.
- To identify the important functionalities such an overlay computer should provide as tools to the programmer, and to develop, rigorously analyze and experimentally validate algorithmic methods that can make these functionalities efficient, scalable, fault-tolerant, and transparent to heterogeneity.
- To provide improved methods for communication and computing among wireless and possibly mobile nodes so that they can transparently become part of larger Internet-based overlay computers.
- To implement a set of functionalities, integrate them under a common software platform in order to provide the basic primitives of an overlay computer, as well as build sample services on this overlay computer, thus providing a proof-of-concept for our theoretical results.

8.2.5. EC-GIN

Participants: Pascale Vicat-Blanc Primet, Paulo Gonçalves, Patrick Loiseau, Damien Ancelin, Sébastien Soudan, Romaric Guiller, Ludovic Hablot.

EC-GIN (Europe-China Grid InterNetworking) is an European STREP project started in November 1st 2006. The university of Innsbrück (Austria) is the prime contractor.
The Internet communication infrastructure (the TCP/IP protocol stack) is designed for broad use; as such, it does not take the specific characteristics of Grid applications into account. This one-size-fits-all approach works for a number of application domains, however, it is far from being optimal - general network mechanisms, while useful for the Grid, cannot be as efficient as customised solutions. While the Grid is slowly emerging, its network infrastructure is still in its infancy. Thus, based on a number of properties that make Grids unique from the network perspective, the project EC-GIN will develop tailored network technology in dedicated support of Grid applications. These technical solutions will be supplemented with a secure and incentive-based Grid Services network traffic management system, which will balance the conflicting performance demand and the economic use of resources in the network and within the Grid.
By collaboration between European and Chinese partners, EC-GIN parallels previous efforts for real-time multimedia transmission across the Internet: much like the Grid, these applications have special network requirements and show a special behaviour from the network perspective. However, while research into network support for multimedia applications has flourished, leading to a large number of standard protocols and mechanisms, the research community has neglected network support for Grid computing up to now. By filling this gap and appropriately exploiting / disseminating the project results, EC-GIN will, therefore, cause a "snowball effect" in the European and Chinese networking and Grid computing research communities.
Technically, EC-GIN will make the Grid work, operate, and communicate better. By appropriately utilising the underlying network, Grid resources in general will be used more efficiently and amplify the impact of Grid computing on the society and economy of Europe and China. Year funding: 100Keuros

8.3. International actions

8.3.1. NEGST: JSPT-CNRS

Participants: Olivier Glück, Sébastien Soudan, Romaric Guillier, Ludovic Hablot, Laurent Lefèvre, Pascale Vicat-Blanc Primet, Paulo Gonçalves, Patrick Loiseau, Jean-Christophe Mignot.

The objective of this project is to promote the collaborations of Japan and France on grid computing technology. In order to promote the collaborative researches, we consider that this project is organized for the following three parts:

1. Grid interoperability and applications
2. Grid Metrics
3. Instant Grid and virtualization of grid computing resources.

RESO mainly participates to the Grid Metrics topic.
Despite the development of strong technologies in all these domains, many issues are still open about the measurement methodology itself, the emulation or simulation of Grid platforms and the understanding of Grid software stack, application performance, and fault tolerance. The Grid Metrics topics, basically gathers all researches about applications, programming models, libraries, runtimes, operating systems and network evaluation, either in synthetic environment (emulators and simulators) or real environment (real network and Grids).

8.3.2. AIST Grid Technology Research Center: GridNet-FJ associated team

Participants: Pascale Vicat-Blanc Primet, Olivier Gluck, Ludovic Hablot, Sébastien Soudan, Romaric Guillier, Olivier Gluck, Paulo Gonçalves, Patrick Loiseau.

Since 2007, RESO is pursuing its collaboration with AIST through the Gridnet-FJ associated team. We followed and even increased our working program on four parts: 1) High speed transport protocol over very high speed links, 2) Bandwidth allocation and control in Grids, 3) Optimisation of MPI communications in Grids, 4) Co-design of GtrcNET-packet capture functionality.

On point 1) with the high speed testbed for protocol evaluation we have deployed within Grid5000 and which integrates GtrcNET1 and GtrcNET10, we pursued our work on TCP variants comparison. We highlight the problem of congestion level which makes TCP behave very strangely (long TCP stops) and the problem of congesting reverse traffic. During our visit to AIST, we had long discussions on the TCP stop problem. This issue has been now solved and a patch to LINUX TCP stack has been posted; We also work together on the INRIA HSTTS (High Speed Transport Test Suite) and gathered very interesting and constructives remarks from AIST colleagues.

The collaboration between AIST GTRC team and INRIA RESO team on the point 2) aims at studying how BDTS, a scheduled data transfer service could benefit this flexibility offered by advance provisioning of some network path and to develop a service which use the interface.

AIST GTRC is collaborating with Pr Ishikawa team at University of Tokyo on GridMPI implementation (point 3). The AIST develop GridMPI, an MPI implementation designed for grids, that uses a similar system of relays as we do in our MPI5000 layer. This architecture has two goals: transmission of messages from a private cluster to a public cluster and also allowing the transfer from a Myrinet cluster to TCP on the long distance link. Their implementation is able to manage many relays to forward datas on long distance links. Their objective is different but the involved mechanisms are similar to the MPI5000 layer. The AIST planned to test their relays on Grid5000 to have a real network platform instead of emulating latency as they are doing for the moment. These experiments are realised jointly with the RESO team.

During the stay of Yuestu Kodama and Tomohiro Kudoh at ENS, the INRIA RESO team and the AIST GTRC team design and develop the Metroflux system for 10Gb/s speed and deploy 10 of such equipment within Grid5000/ALADDIN.

8.3.3. Collaboration with University of Lisbon, Portugal

Participant: Paulo Gonçalves.

P. Gonçalves is co-advising the PhD program of Hugo Carrão from ISEGI, University of Lisbon. A grant from the “Programme Actions Universitaires Intégrées Luso-Françaises" supports our collaboration (ends in january 2009). H. Carrão PhD defense is programmed in january 27, 2010.

8.4. Visitors

8.4.1. Collaboration with University of Sevilla, Spain

Participants: Laurent Lefèvre, Anne-Cécile Orgerie.

RESO has hosted one PhD students from University of Sevilla for short term periods during the year 2009 : Alejandro Fernandez (September - November 2009) to work on the prediction models for energy efficiency in large scale distributed systems.

9. Dissemination

9.1. Conference organisation, editors for special issues

9.1.1. Editorial Boards

- *Computer Communications*, Elsevier, I. Guérin Lassous
- *Ad Hoc Networks*, Elsevier, I. Guérin Lassous
- *Discrete Mathematics and Theoretical Computer Science*, I. Guérin Lassous
- *Performance Evaluation*, Elsevier, I. Guérin Lassous, Guest Editor of Special issue on on "Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks".
- *Future Generation Computer Systems (FGCS)*, Elsevier, Pascale Vicat-Blanc, Guest Editor of Special issue on "High Speed Networks for Grid Applications"
- *Annals of Telecoms* P. Vicat-Blanc Primet, Guest Editor of a special issue on "Grid, Cloud and Utility computing".
- *Lecture Notes in Computer Science* P. Vicat-Blanc Primet, Guest Editor
- *"Scaling, Fractals and Wavelets"*, John Whiley Ed., P. Gonçalves, co-editor.

9.1.2. Chairing and Organisation of Conferences and Workshops

- *Parco 2009*, Laurent Lefèvre
- *ACM PE-WASUN 2009*, I. Guérin Lassous, PC co-chair
- *ACM MobiHoc 2009*, I. Guérin Lassous, Poster chair
- *IEEE BroadNet 2009*, P. Vicat-Blanc Primet, PC co-chair
- *CSA 2009*, Laurent Lefèvre, Track co-chair
- *GADA2009*, Laurent Lefèvre, Program Comittee Co-Chair
- *E2GC2*, Laurent Lefèvre, Workshop co-chair and organizer
- *Parco 2009*, Laurent Lefèvre, Local Co-Organizer
- *IEEE HPCC-09*, Laurent Lefèvre, General co-chair
- *HPPAC 2009*, Laurent Lefèvre, Workshop Co-Chair

9.1.3. Program committee members

- *ITC*, P. Vicat-Blanc Primet, 2009
- *IEEE/ACM CCGrid*, P. Vicat-Blanc Primet, 2009, 2008
- *ITCSS*, P. Vicat-Blanc Primet, 2009
- *CFIP*, P. Vicat-Blanc Primet, 2009, 2008, 2007
- *ISWCS*, I. Guérin Lassous, 2009

- *MedHocNet*, I. Guérin Lassous, 2009, 2008, 2006
- *HotMesh*, I. Guérin Lassous, 2009
- *VTC-Spring*, I. Guérin Lassous, 2009
- *PerSens*, I. Guérin Lassous, 2009, 2008, 2007, 2006
- *ICDCN*, I. Guérin Lassous, 2009, 2008
- *IEEE VTC*, Eric Fleury, 2009
- *IEEE PERCOM*, Eric Fleury, 2009
- *IEEE ICC*, Eric Fleury, 2009
- *COMSNET*, Eric Fleury, 2009
- *IEEE/IFIP EUC*, Eric Fleury, 2009
- *PDCAT*, Laurent Lefèvre, 2009
- *MGC*, Laurent Lefèvre, 2009, 2008
- *IEEE SuperComputing*, Laurent Lefèvre, 2009, 2006
- *NSS*, Laurent Lefèvre, 2009
- *CFSE*, Laurent Lefèvre, 2009
- *Renpar*, Laurent Lefèvre, 2009
- *ICCN*, Laurent Lefèvre, 2009
- *IEEE/ACM HPDC*, Laurent Lefèvre, 2009, 2008, 2007
- *ICCS*, Laurent Lefèvre, 2009, 2008
- *HotP2P*, Laurent Lefèvre, 2009, 2008, 2007, 2006
- *IEEE/ACM CCGrid*, Laurent Lefèvre, 2009
- *DAGRES*, Laurent Lefèvre, 2009
- *AusGrid*, Laurent Lefèvre, 2009, 2008

9.1.4. Participation in steering committees

- IEEE/ACM CCGrid conference, Laurent Lefèvre, since 2004
- ACM GridNets, Pascale Vicat-Blanc Primet, since 2006
- PFLDNET workshop series, P. Vicat-Blanc Primet, since 2005
- ICPS, Laurent Lefèvre since 2006

9.1.5. International expertise

- *PhD examining boards*, P. Vicat-Blanc Primet : F. Dijkstra (University of Amsterdam, reviewer 2009)
- *PhD examining boards*, Laurent Lefèvre : Lakshmi Priya (Anna University, Chennai, India, reviewer, 2009)
- *PhD examining boards*, I. Guérin Lassous : Sandrine Calomme (University of Liege, Belgium, reviewer, 2009), Lars Landmark (Norwegian University of Science and Technology, Norway, first opponent, 2009)

9.1.6. National expertise

- *PhD examining boards*, Isabelle Guérin Lassous, 2009: Despoina Triantafyllidou (University Paris-Sud, reviewer), Fahrid Munir (EURECOM, examinator), Fadila Khadar (University Lille 1, reviewer), Husnain Mansoor Ali (University Paris-sud 11, examinator).

- *HdR examining board*, Isabelle Guérin Lassous, 2009 : Mohamed Senouci (Orange Labs, reviewer).
- *PhD examining boards*, P. Vicat-Blanc Primet 2009: Lila Boukhatem (University of Orsay, reviewer, 2009), Nicolas Van Wambeke (Laas, Toulouse, reviewer), Ala Resmerita (LRI, Orsay, reviewer), Carlos Barrios Hernandez (LIG, Grenoble, reviewer)
- *PhD examining boards*, Laurent Lefèvre 2009: Anthony Mouraud (University Antilles Guyane, examinator)
- *PhD examining boards*, Olivier Glück 2009: François Trahay (University Bordeaux 1, examinator)

9.1.7. Public Dissemination

- Interstice, Podcast, "Very High Speed Networks", P. Vicat-Blanc Primet, May 2009
- Usine Nouvelle, "Construire l'Internet de demain", Interview de P. Vicat-Blanc Primet, Octobre 2009
- INEDIT : the Newsletter of INRIA, "Towards green computing platforms - vers des plateformes de calcul vertes", Laurent Lefèvre, May 2009
- INRIA Booth at Supercomputing Conference(SC), Laurent Lefèvre, 2007, 2009.

9.2. Graduate teaching

- **since 2006** I. Guérin Lassous
  Multimédia and Quality of Service Master 2 SIR (Professional) / RTS (Research) (University Claude Bernard Lyon I), lecture 18h, others 12h.
- **since 2006** I. Guérin Lassous
  Networking
  Master 2 CCI (Professional)(University Claude Bernard Lyon I), lecture 18h, others 12h.
- **since 2006** I. Guérin Lassous
  Autonomic Computing
  Master 2 RTS (Research) University Claude Bernard Lyon I), lecture 15h.
- **since 2004** O.Glück
  Client/Server Model, Internet Applications, Network and System Administration.
  Master 2 SIR (University Claude Bernard Lyon 1), lecture 30h, others 30h.
- **since 2004** JP.Gelas
  Long Distance networks ; Networks and Transport Protocols ; QoS and Multimedia ; Initiation to Java ; Local Area Networks .
  Master 2 SIR and CCI (University Claude Bernard Lyon 1), lecture 30h, others 40h.
- **since 2005** JP.Gelas
  Long distance networks ; Networks and Transport Protocols ; Routing ; Advanced Java and Web services.
  Master 2 SIR (Université Claude Bernard Lyon 1), lecture 45h, others 45h.
- **since 2007** JP.Gelas
  Embedded System and Software.
  Master 2 SIR, TI, Image and App (Université Claude Bernard Lyon 1), lecture 30h, others 30h.
- **since 2008** JP.Gelas
  Introduction to System, Computer Networks and Client/Server architecture.
  Master 2 CCI (Université Claude Bernard Lyon 1), lecture 20h, others 30h.
- **since 2009** T.Begin
  Computer Networks.
  Master 1 (Université Claude Bernard Lyon 1), lecture 12h, others 40h.

- **since 2004** T.Begin
  Client/Server Model, Internet Applications, Network and System Administration.
  Master 2 SIR/TIW (University Claude Bernard Lyon 1), practical work 32h.
- **2007, 2009** P. Gonçalves
  Models for Traffic.
  Master 2, Dept. Informatique fondamentale, ENS Lyon. Research class (24h).

9.3. Miscelleneous teaching

- **since 2006**: I. Guérin Lassous
  Ad Hoc Networks Master 1 (University Claude Bernard Lyon I), lecture 6h, others 6h.
- **since 2004**: O. Glück
  Computer Networks.
  Licence Informatique, (University Claude Bernard Lyon 1), lecture 30h, others 30h.
- L. Lefèvre
  is responsible of training periods for Research Master in ENS-Lyon
- **since 2007**: JP. Gelas
  Long Distance and High Performance Network.
  Graduate students of the "Institut de la Francophonie pour l'Informatique" in HanoÃ¯, Vietnam, 60h lectures.
- **since 2009**: T. Begin
  Computer Networks.
  Licence Informatique, (University Claude Bernard Lyon 1), tutorials and practical works 34h.
- **since 2009**: P. Gonçalves is responsible for the axis "Models and Optimization for Emergent infrastructures" of the ENS Lyon Computer Science Master (Informatique fondamentale)

9.4. Animation of the scientific community

- Pascale Vicat-Blanc
  - the INRIA scientific leader of ADR Semantic Netwokring of the INRIA Bell Labs common laboratory
  - is member of the "Networks" expert committee of the CNRS.
  - is within the Grid5000 project and ADT ALADDIN, member of the steering committee and co-leader of the Grid5000@Lyon site.
  - is leading the ANR CIS HIPCAL project.
  - is leading the INRIA team within the CARRIOCAS System@tic project.
  - is leading the INRIA team within the european EC-GIN project.
  - is leading the LIP team of the ANR (blanc) IGTMD project.
- Isabelle Guérin Lassous is:
  - member of the CNRS TAROT action (Techniques Algorithmiques, Réseaux et d'Optimisation pour les Télécommunications);
  - the INRIA scientific leader of the european project AEOLUS (Algorithmic Principles for Building Efficient Overlay Computers);
  - member of the new INRIA - Bell Labs common research laboratory (INRIA scientific leader of the WP4 - Mechanisms for QoS control and management of flows in the Semantic Networking research activity).

- Laurent Lefèvre is :
  - leading the INRIA ARC GREEN-NET on “Power aware software frameworks for high performance data transport and computing in large scale distributed systems” which involved 4 partners : INRIA RESO, INRIA MESCAL (Grenoble), IRIT (Toulouse), Virginia Tech (USA).
  - leading the WorkPackage 5 on “Service Deployment” of the FP7 STREP Project “Autonomic Internet”
  - the INRIA representative of the OGF-Europe project
  - Management Committee Member and French representative of the COST Action IC0804 on Energy efficiency in large scale distributed systems.

9.5. Participation in boards of examiners and committees

- Pascale Vicat-Blanc : president of the hearing committee of INRIA Rhône-Alpes;
- Isabelle Guérin Lassous is member of:
  - CNU, section 27;
  - the specialists committee (section 27) of the UJF (Université Joseph Fourier) - Grenoble;
  - selection committee of the Ecole Nationale d’Administration (ENA).
- Olivier Glück is a member of
  - two PhD examining boards: François Trahay (University Bordeaux 1 - examinator), Ludovic Hablot (ENS Lyon - examinator);
  - the 27ème section selection committee of University Claude Bernard Lyon 1;
  - the “conseil du département Informatique” of University Claude Bernard Lyon 1;
  - the “conseil de l’UFR Faculté des Sciences et Technologies” of University Claude Bernard Lyon 1;
  - the “conseil des Etudes et de la Vie Universitaire” of University Claude Bernard Lyon 1.
- Laurent Lefèvre is member of :
  - CNU Section 27;
- Jean-Patrick Gelas is :
  - co-manager of the professional master CCI (SIRR) (University Claude Bernard Lyon 1)
  - selection committee member (section 27) of the UCBL (University Claude Bernard Lyon 1)

9.6. Seminars, invited talks

- Isabelle Guérin Lassous gave:
  - a seminar of one week on "quality of service in multihop wireless networks" at UPC, Barcelona, Spain, June 2009.
- Laurent Lefèvre has been invited to give the following talks :
  - "Why hunting watts in large scale distributed systems ? The GREEN-* approaches !", Laurent Lefèvre, Opening Keynote Talk, Renpar 2009 : Rencontres francophones du Parallélisme, Toulouse, France, September 2009

- "A Service Enabler Infrastructure for the Future Internet", Laurent Lefèvre and Abderhaman Cheniour, Joint EMANICS, AutoI, Self-Net Workshop on Autonomic Management, London UK, April 2009
- "Energy Efficiency issues for large scale distributed systems : the GREEN-NET initiative", Laurent Lefèvre, OGF 25 : Open Grid Forum during "OGF-EU: Using IT to reduce Carbon Emissions and Delivering the Potential of Energy Efficient Computing" session, Catania, Italy, March 4, 2009
- "Towards Energy Aware Resource Infrastructure for Large Scale Distributed Systems", Laurent Lefèvre and Anne Cécile Orgerie, University of Melbourne, Australia, January 14, 2009

10. Bibliography

Major publications by the team in recent years

[1] F. BOUHAFS, J.-P. GELAS, L. LEFÈVRE, M. MAIMOUR, C. PHAM, P. VICAT-BLANC PRIMET, B. TOURANCHEAU. *Designing and Evaluating An Active Grid Architecture*, in "The International Journal of Future Generation Computer Systems (FGCS) - Grid Computing: Theory, Methods and Applications", vol. 21, n^{o} 2, February 2005, p. 315-330.

[2] F. CAPPELLO, F. DESPREZ, M. DAYDE, E. JEANNOT, Y. JEGOU, S. LANTERI, N. MELAB, R. NAMYST, P. VICAT-BLANC PRIMET, O. RICHARD, E. CARON, J. LEDUC, G. MORNET. *Grid5000: a nation wide experimental grid testbed*, in "in International Journal on High Performance Computing Applications", 2006.

[3] B. B. CHEN, P. VICAT-BLANC PRIMET. *Supporting bulk data transfers of high-end applications with guaranteed completion time*, in "IEEE ICC2007 International Conference on Computer Communication", IEEE, 2007.

[4] B. GOGLIN, O. GLÜCK, P. VICAT-BLANC PRIMET. *An Efficient Network API for in-Kernel Applications in Clusters*, in "Proceedings of the IEEE International Conference on Cluster Computing, Boston, Massachussets", IEEE Computer Society Press, September 2005.

[5] P. GONÇALVES, R. RIEDI. *Diverging moments and parameter estimation*, in "Journal of American Statistical Association", vol. 100, n^{o} 472, December 2005, p. 1382–1393.

[6] J. LAGANIER, P. VICAT-BLANC PRIMET. *HIPernet: a decentralized security infrastructure for large scale grid environments*, in "6th IEEE/ACM International Conference on Grid Computing (GRID 2005), November 13-14, 2005, Seattle, Washington, USA, Proceedings", IEEE, 2005, p. 140-147.

[7] L. LEFÈVRE, J.-P. GELAS. *Chapter 14 on "High Performance Execution Environments"*, in "Programmable Networks for IP Service Deployment", A. GALIS, S. DENAZIS, C. BROU, C. KLEIN (editors), Artech House Books, UK, may 2004, p. 291-321.

[8] D. LOPEZ PACHECO, C. PHAM, L. LEFÈVRE. *XCP-i : eXplicit Control Protocol for heterogeneous inter-networking of high-speed networks*, in "Globecom 2006, San Francisco, California, USA", November 2006.

Year Publications

Doctoral Dissertations and Habilitation Theses

[9] R. GUILLIER. *Methodologies and Tools for the Evaluation of Transport Protocols in the Context of Highspeed Networks*, ENS-Lyon, Université de Lyon, September 2009, Ph. D. Thesis.

[10] L. HABLOT. *Réseau longue distance et application distribuée dans les grilles de calcul : étude et propositions pour une interaction efficace*, ENS Lyon, Université de Lyon, 2009, Ph. D. Thesis.

[11] P. LOISEAU. *Contributions to the analysis of scaling behavior and quality of service in networks: experimental and theoretical aspects*, École Normale Supérieure de Lyon, December 2009, Ph. D. Thesis.

[12] V. RÉMI. *Partage de la bande passante équitable dans les réseaux ad hoc*, ENS Lyon, INRIA, 2009, Ph. D. Thesis.

[13] S. SOUDAN. *Bandwidth Sharing and Control in High-Speed Networks: Combining Packet- and Circuit-Switching Paradigms*, ENS-Lyon, Université de Lyon, 46, allée d'Italie, 69364 Lyon cedex 07, France, 2009, Ph. D. Thesis.

Articles in International Peer-Reviewed Journal

[14] O. AUDOUIN, D. BARTH, M. GAGNAIRE, C. MOUTON, P. VICAT-BLANC PRIMET, D. RODRIGUES, L. THUAL, D. VERCHÈRE. *CARRIOCAS project: Towards Converged Internet Infrastructures Supporting High Performance Distributed Applications*, in "IEEE/OSA Journal of Lightwave Technology", 2009, accepted.

[15] H. CARRÃO, A. ARAÚJO, P. GONÇALVES, M. CAETANO. *Multitemporal MERIS images for land cover mapping at national scale: the case study of Portugal*, in "International Journal of Remote Sensing", 2009, To appear.

[16] H. CARRÃO, P. GONÇALVES, M. CAETANO. *A nonlinear model for satellite images time series: analysis and prediction of land cover dynamics*, in "IEEE Trans. in Geosciences and Remote Sensing", 2009, To appear.

[17] L. LEFÈVRE, A.-C. ORGERIE. *Designing and Evaluating an Energy Efficient Cloud*, in "Journal of Super-Computing", December 2009.

[18] L. LEFÈVRE, A.-C. ORGERIE. *Towards Energy Aware Reservation Infrastructure for Large-Scale Experimental Distributed Systems*, in "Parallel Processing Letters - Special Issue on Clusters and Computational Grids for Scientific Computing", vol. 19, n^{o} 3, September 2009, p. 419-433.

[19] P. LOISEAU, P. GONÇALVES, G. DEWAELE, P. BORGNAT, P. ABRY, P. VICAT-BLANC PRIMET. *Investigating self-similarity and heavy-tailed distributions on a large scale experimental facility*, in "IEEE/ACM Transactions on Networking", December 2009, http://perso.ens-lyon.fr/patrick.loiseau/articles/Invest_SS_and_HT.pdf, to appear.

[20] P. NEIRA AYUSO, R. GASCA, L. LEFÈVRE. *Demystifying Cluster-Based Fault-Tolerant Firewalls*, in "IEEE Internet Computing : Special Issue on Unwanted Traffic", vol. 13, n^{o} 6, November 2009, p. 30-37.

[21] P. SPINNATO, P. VICAT-BLANC PRIMET, C. EDWARDS, M. WELZL. *Editorial Special Section on Networks for Grid Applications*, in "International Journal on Future Generation Computer Systems", april 2009.

[22] P. VICAT-BLANC PRIMET, S. SOUDAN, D. VERCHERE. *Virtualizing and scheduling optical network infrastructure for emerging IT services*, in "Optical Networks for the Future Internet (special issue of Journal of Optical Communications and Networking (JOCN))", vol. 1, n^{o} 2, 2009, p. A121–A132, http://jocn.osa.org/abstract.cfm?URI=JOCN-1-2-A121.

International Peer-Reviewed Conference/Proceedings

[23] A. AGAPI, S. SOUDAN, M. PASIN, P. VICAT-BLANC PRIMET, T. KIELMANN. *Optimizing deadline-driven bulk data transfers in overlay networks*, in "ICCCN 2009 Track on Pervasive Computing and Grid Networking (PCGN), San Francisco, USA", 2009.

[24] F. ANHALT, P. VICAT-BLANC PRIMET. *Analysis and experimental evaluation of data plane virtualization with Xen*, in "ICNS 09 : International Conference on Networking and Services, Valencia, Spain", April 2009, http://doi.ieeecomputersociety.org/10.1109/ICNS.2009.77.

[25] N. AYARI, D. BARBARON, L. LEFÈVRE. *Evaluating Session Aware Admission Control Strategies for Improving the Profitability of Service Providers*, in "The 3rd IEEE Workshop on Enabling the Future Service-Oriented Internet: Towards Socially-Aware Networks - Held in conjunction with IEEE GLOBECOM 2009, Honolulu, USA", December 2009.

[26] S. BAYKUT, P. GONÇALVES, P.-H. LUPPI, P. ABRY, E. PEREIRA DE SOUZA NETO, D. GERVASONI. *EMD-based analysis of rat EEG data for sleep state classification*, in "Biosignals", Springer, January 2009.

[27] G. DA-COSTA, J.-P. GELAS, Y. GEORGIOU, L. LEFÈVRE, A.-C. ORGERIE, J.-M. PIERSON, O. RICHARD, K. SHARMA. *The GREEN-NET Framework: Energy Efficiency in Large Scale Distributed Systems*, in "HPPAC 2009 : High Performance Power Aware Computing Workshop in conjunction with IPDPS 2009, Rome, Italy", May 2009.

[28] D. M. DIVAKARAN, E. ALTMAN, G. POST, L. NOIRIE, P. VICAT-BLANC PRIMET. *Analysis of the effects of XLFrames in a network*, in "IFIP/TC6 NETWORKING 2009", May 2009, p. 364–377, http://www.springerlink.com/content/9107653164ng8541.

[29] D. M. DIVAKARAN, E. ALTMAN, G. POST, L. NOIRIE, P. VICAT-BLANC PRIMET. *From Packets to XLFrames: Sand and Rocks for Transfer of Mice and Elephants*, in "IEEE INFOCOM 2009 Workshop on High-Speed Networks, Rio di Janerio, Brazil", Apr 2009, p. 1–6, http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5072149&isnumber=5072090.

[30] E. DRAMITINOS, I. GUÉRIN LASSOUS. *A Bandwidth Allocation Mechanism for 4G*, in "European Wireless Technology Conference, Roma, Italy", September 2009.

[31] E. DRAMITINOS, I. GUÉRIN LASSOUS, R. VANNIER. *A Performance Evaluation Framework for Fair Solutions in Ad Hoc Networks*, in "12-th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM), Tenerife, Canary Islands, Spain", October 2009.

[32] A. GALIS, S. DENAZIS, A. BASSI, P. GIACOMIN, A. BERL, A. FISCHER, H. DE MEER, J. SRASSNER, S. DAVY, D. MACEDO, G. PUJOLLE, J. LOYOLA, J. SERRAT, L. LEFÈVRE, A. CHENIOUR. *Management*

Architecture and Systems for Future Internet Networks, in "FIA Book : "Towards the Future Internet - A European Research Perspective", Prague", IOS Press, May 2009, p. 112-122, ISBN 978-1-60750-007-0.

[33] R. GUILLIER, P. VICAT-BLANC PRIMET. *A User-Oriented Test Suite for Transport Protocols Comparison in DataGrid Context*, in "ICOIN 2009", January 2009, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=4897301&isnumber=4897247.

[34] I. GUÉRIN LASSOUS, T. RAZAFINDRALAMBO, R. VANNIER. *Evaluation d'un protocole de régulation de débit dans les réseaux sans fil multisauts*, in "Conférence Francophone sur l'Ingénierie des Protocoles, Strasbourg, France", October 2009.

[35] G. KOSLOVSKI, T. TRUONG HUU, J. MONTAGNAT, P. VICAT-BLANC PRIMET. *Executing distributed applications on virtualized infrastructures specified with the VXDL language and managed by the HIPerNET framework*, in "First International Conference on Cloud Computing (CLOUDCOMP 2009), Munich, Germany", October 2009, http://www.cloudcomp.eu/.

[36] P. LOISEAU, P. GONÇALVES, S. GIRARD, F. FORBES, P. VICAT-BLANC PRIMET. *Maximum Likelihood Estimation of the Flow Size Distribution Tail Index from Sampled Packet Data*, in "ACM Sigmetrics", June 2009, http://portal.acm.org/citation.cfm?id=1555349.1555380.

[37] P. LOISEAU, P. GONÇALVES, R. GUILLIER, M. IMBERT, Y. KODAMA, P. VICAT-BLANC PRIMET. *Metroflux: A high performance system for analyzing flow at very fine-grain*, in "TridentCom", April 2009.

[38] K. MUNIR, P. VICAT-BLANC PRIMET, M. WELZL. *Grid Network Dimensioning by Modeling the Deadline Constrained Bulk Data Transfers*, in "11th IEEE International Conference on High Performance Computing and Communications (HPCC-09), Seoul, Korea", June 2009.

[39] L. NUSSBAUM, F. ANHALT, O. MORNARD, J.-P. GELAS. *Linux-based virtualization for HPC clusters*, in "Linux Symposium 2009", July 2009.

[40] L. NUSSBAUM. *Rebuilding Debian using Distributed Computing*, in "6th International Workshop on Challenges of Large Applications in Distributed Environments (CLADE 2009), held in conjunction with the International ACM Symposium on High Performance Distributed Computing (HPDC 2009)", June 2009.

[41] A.-C. ORGERIE, L. LEFÈVRE. *When Clouds become Green: the Green Open Cloud Architecture*, in "Parco2009 : International Conference on Parallel Computing, Lyon, France", September 2009.

[42] T. RAZAFINDRALAMBO, I. GUÉRIN LASSOUS. *SBA: a Simple Backoff Algorithm for Wireless Ad Hoc Networks*, in "IFIP Networking, Aachen, Germany", May 2009.

[43] S. SOUDAN, D. M. DIVAKARAN, E. ALTMAN, P. VICAT-BLANC PRIMET. *Equilibrium in size-based scheduling systems*, in "16th International Conference on Analytical and Stochastic Modelling Techniques and Applications, Madrid, Spain", Jun 2009, p. 234–248.

[44] S. SOUDAN, P. VICAT-BLANC PRIMET. *Mixing Malleable and Rigid Bandwidth Requests for Optimizing Network Provisioning*, in "21st International Teletraffic Congress, Paris, France", sept 2009.

[45] P. VICAT-BLANC PRIMET, F. ANHALT, G. KOSLOVSKI. *Exploring the virtual infrastructure service concept in Grid'5000*, in "20th ITC Specialist Seminar on Network Virtualization, Hoi An, Vietnam", May 2009, http://www.itcspecialistseminar.com/paper/itcss09_Primet.pdf.

[46] P. VICAT-BLANC PRIMET, J.-P. GELAS, O. MORNARD, G. KOSLOVSKI, V. ROCA, L. GIRAUD, J. MONTAGNAT, T. T. HUU. *A scalable security model for enabling Dynamic Virtual Private Execution Infrastructures on the Internet*, in "IEEE/ACM International Conference on Cluster Computing and the Grid (CCGrid2009), Shanghai", May 2009, http://portal.acm.org/citation.cfm?id=1577923.

National Peer-Reviewed Conference/Proceedings

[47] F. ANHALT, G. KOSLOVSKI, M. PASIN, J.-P. GELAS, P. VICAT-BLANC PRIMET. *Les Infrastructures Virtuelles à la demande pour un usage flexible de l'Internet*, in "JDIR 09 : Journees Doctorales en Informatique et Reseaux, Belfort, France", February 2009, http://jdir.utbm.fr/articles/anhalt.pdf.

[48] A.-C. ORGERIE, L. LEFÈVRE, J.-P. GELAS. *Economies d'Energie dans les Systèmes Distribués à Grande Echelle : l'Approche EARI*, in "JDIR 09 : Journèes Doctorales en Informatique et Réseaux, Belfort, France", February 2009.

[49] C. SARR, S. KHALFALLAH, I. GUÉRIN LASSOUS. *Gestion dynamique de la bande passante dans les réseaux ad hoc multi-sauts*, in "9es Journées Doctorales Informatique et Réseau (JDIR), Belfort, France", January 2009.

Scientific Books (or Scientific Book chapters)

[50] A. BOUKERCHE, I. GUÉRIN LASSOUS (editors). *Sixth ACM International Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks (PE-WASUN)*, ACM, ACM, Canary Islands, Spain, October 2009.

[51] P. ABRY, P. GONÇALVES, J. LÉVY VÉHEL. *Scaling, Fractals and wavelets*, Digital signal and image processing series, ISTE – John Wiley & Sons, Inc., London (UK) and Hoboken (NJ, USA), 2009.

[52] C. CHAUDET, I. GUÉRIN LASSOUS. *Principes et protocoles d'accès au médium*, in "Réseaux de capteurs", ISTE/Wiley, 2009.

[53] Z. HUANG, Z. XU, L. LEFÈVRE, H. SHEN, J. HINE, Y. PAN. *Special Issue on "Emerging Research in Parallel and Distributed Computing" - Journal of Supercomputing*, Springer, December 2009.

[54] L. LEFÈVRE, J.-M. PIERSON. *Special theme ERCIM News : Towards Green ICT*, vol. 79, ERCIM, October 2009.

Books or Proceedings Editing

[55] P. SPINNATO, P. VICAT-BLANC PRIMET, C. EDWARDS, M. WELZL (editors). *Special Section on Networks for Grid Applications*, Elsevier, april 2009, International Journal on Future Generation Computer Systems.

[56] P. VICAT-BLANC PRIMET, T. KUDOH, J. MAMBRETTI (editors). *Networks for Grid Applications*, Springer, Beijing, China, june 2009.

Research Reports

[57] D. M. DIVAKARAN, G. CAROFIGLIO, E. ALTMAN, P. PRIMET. *A Flow Scheduler Architecture*, n^{o} RR-7133, INRIA, 2009, http://hal.inria.fr/inria-00438594/en/, Research Report.

[58] D. M. DIVAKARAN, S. SOUDAN, P. VICAT-BLANC PRIMET, E. ALTMAN. *A survey on core switch designs and algorithms*, n^{o} RR-6942, INRIA, 2009, http://hal.inria.fr/inria-00388943/en/, Research Report.

[59] E. DRAMITINOS, R. VANNIER, I. GUÉRIN LASSOUS. *A Utility-based Framework for Assessing Fairness Schemes in Ad-Hoc Networks*, INRIA, 2009, http://hal.inria.fr/inria-00360848/en/, Technical report.

[60] G. FEDAK, J.-P. GELAS, T. HÉRAULT, V. INIESTA, D. KONDO, L. LEFÈVRE, P. MALECOT, L. NUSSBAUM, A. REZMERITA, O. RICHARD. *DSL-Lab: a Platform to Experiment on Domestic Broadband Internet*, n^{o} RR-7024, INRIA, 2009, http://hal.inria.fr/inria-00424936/en/, Research Report.

[61] R. GUILLIER, S. SOUDAN, P. VICAT-BLANC PRIMET. *UDT and TCP without Congestion Control for Profile Pursuit*, n^{o} 6874, INRIA, 03 2009, http://hal.inria.fr/inria-00367160/fr/, Also available as LIP Research Report RR2009-10, Research Report.

[62] L. HABLOT, O. GLÜCK, J.-C. MIGNOT, R. GUILLIER, S. SOUDAN, P. VICAT-BLANC PRIMET. *Interaction between MPI and TCP in grids*, n^{o} 6945, INRIA, 06 2009, http://hal.inria.fr/inria-00389836/, Research Report.

[63] P. LOISEAU, P. GONÇALVES, P. VICAT-BLANC PRIMET. *Impact of the Correlation between Flow Rates and Durations on the Large-Scale Properties of Aggregate Network Traffic*, n^{o} 7100, INRIA, November 2009, Technical report.

[64] A.-C. ORGERIE, L. LEFÈVRE. *A year in the life of a large-scale experimental distributed system: usage of the Grid'5000 platform in 2007*, n^{o} 6965, INRIA, April 2009, http://hal.inria.fr/inria-00400684/en/, Research Report.

[65] S. SOUDAN, D. M. DIVAKARAN, E. ALTMAN, P. VICAT-BLANC PRIMET. *Equilibrium in size-based scheduling systems*, n^{o} 6888, INRIA, 03 2009, http://hal.inria.fr/inria-00371391/en/, Also available as LIP Research Report RR2009-11, Research Report.

[66] S. SOUDAN, D. M. DIVAKARAN, E. ALTMAN, P. VICAT-BLANC PRIMET. *Extending Routing Games to Flows over Time*, n^{o} 6931, INRIA, 05 2009, Research Report.

Patents and standards

[67] R. GUILLIER, P. VICAT-BLANC PRIMET. *PATHNIF. INRIA Patent*, 2009.

Other Publications

[68] W. CHAI, L. MAMATAS, A. GALIS, J. LOYOLA, J. SERRAT, A. FISCHER, A. PALER, Y. AL-HAZMI, A. BERL, H. DE MEER, A. CHENIOUR, L. LEFÈVRE, S. DAVY, D. MULDOWNEY, G. KOUMOUTSOS, A. BASSI, Z. MOVAHEDI. *Open Source for Future Internet Systems*, November 2009, FIA2009 : Future Internet Assembly in Stockholm - Poster.

[69] G. DA-COSTA, J.-P. GELAS, Y. GEORGIOU, L. LEFÈVRE, A.-C. ORGERIE, J.-M. PIERSON, O. RICHARD. *The GREEN-NET approach for supporting energy efficient solutions in Grids*, September 2009, Short paper and Poster in Renpar 2009 : French Meeting in Parallelism.

[70] I. GUÉRIN LASSOUS. *Standard pour les réseaux sans fil : IEEE 802.11*, n^{o} 7375, 2009, Journal Techniques de l'Ingénieur.

[71] L. LEFÈVRE, A. CHENIOUR. *A Service Enabler Infrastructure for the Future Internet*, April 2009, Joint EMANICS, AutoI, Self-Net Workshop on Autonomic Management, London UK.

[72] L. LEFÈVRE. *Towards green computing platforms - Vers des plateformes de calcul vertes*, May 2009.

[73] A.-C. ORGERIE, L. LEFÈVRE. *Greening the Clouds !*, June 2009, Poster during Rescom 2009 Summer School, La Palmyre, France.

[74] A.-C. ORGERIE, L. LEFÈVRE. *Towards a Green Grid5000*, April 2009, Best presentation award of the Grid5000 school.

References in notes

[75] N. AYARI, D. BARBARON, L. LEFÈVRE. *Procédés de gestion de sessions multi-flux. France Telecom R&D Patent*, June 2007.

[76] N. AYARI, D. BARBARON, L. LEFÈVRE, P. VICAT-BLANC PRIMET. *Implementation of an Active Replication based Framework for Highly Available Services*, September 2007, NetFilter Workshop 2007, Karlsruhe, Germany.

[77] N. AYARI, D. BARBARON, L. LEFÈVRE, P. VICAT-BLANC PRIMET. *SARA: A Session Aware Infrastructure for High Performance Next Generation Cluster-based Servers*, in "ATNAC 2007 : Australasian Telecommunication Networks and Applications Conference, Christchurch, New Zealand", December 2007.

[78] N. AYARI, D. BARBARON, L. LEFÈVRE, P. VICAT-BLANC PRIMET. *Session Awareness issues for next-generation cluster-based network load balancing frameworks*, in "AICCSA07 : ACS/IEEE International Conference on Computer Systems and Applications, Amman, Jordan", May 2007, p. 180-186.

[79] N. AYARI, D. BARBARON, L. LEFÈVRE, P. VICAT-BLANC PRIMET. *T2CP-AR: A system for Transparent TCP Active Replication*, in "AINA-07 : The IEEE 21st International Conference on Advanced Information Networking and Applications, Niagara Falls, Canada", May 2007, p. 648-655.

[80] C. ESTAN, G. VARGHESE. *New directions in traffic measurement and accounting*, in "Proceedings of the SIGCOMM conference", vol. 32, n^{o} 4, 2002, p. 323–336.

[81] S. FLOYD, V. JACOBSON. *Link-sharing and Resource Management Models for Packet Networks*, in "IEEE/ACM Transaction on Networking", 4, vol. 3, August 1995.

[82] I. FOSTER, C. KESSELMAN. *The Grid : Blueprint for a new Computing Infrastructure*, in "Morgan Kaufmann Publishers Inc.", 1998.

[83] O. GREÉMILLET, P. GONÇALVES, P. VICAT-BLANC PRIMET, A. DUPAS. *Traffic classification techniques supporting semantic networks*, in "1st Int. Wireless Comm. and Mobile Comp. Conf., TRaffic Analysis and Classification W., Caen (France)", Jun-jul 2010, Submitted.

[84] L. HABLOT, O. GLÜCK, J.-C. MIGNOT, P. VICAT-BLANC PRIMET. *Etude d'implémentations MPI dans une grille de calcul*, in "Actes de Renpar'08", Février 2008.

[85] D. LOPEZ PACHECO, L. LEFÈVRE, C. PHAM. *XCP-i : eXplicit Control Protocol pour l'interconnexion de réseaux haut-débit hétérogènes*, n^{o} 6385, INRIA, December 2007, http://hal.inria.fr/inria-00195634, Also available as LIP Research Report RR2007-47, Research Report.

[86] A. W. MOORE, D. ZUEV. *Internet traffic classification using bayesian analysis techniques*, in "ACM SIGMETRICS", 2005, p. 50–60.

[87] V. SANDER. *Networking issues of GRID Infrastructures*, in "GRID Working Draft of the GRID High-Performance Networking Research Group, Global GRID Forum", 2003.

[88] V. SANDER, F. TRAVOSTINO, J. CROWCROFT, P. VICAT-BLANC PRIMET, C. PHAM. *Networking Issues of Grid Infrastructures*, Open Grid Forum, october 2004, http://forge.gridforum.org/projects/ghpn-rg/, Technical report.

[89] D. SIMEONIDOU. *Optical Network Infrastructure for Grid*, in "Grid Working Draft of the Grid High-Performance Networking Research Group, Global GRID Forum", 2003.