Energy efficiency in large scale distributed systems

RESO, Lyon

Rhône-Alpes

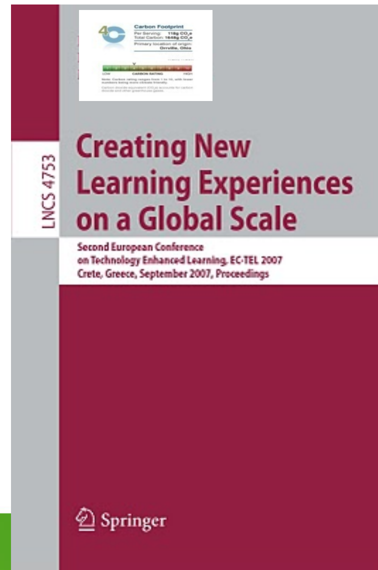
2012, March 22

Energy : 1st challenge for large scale systems (datacenter, grids, clouds, internet)

Future exascale platforms -> systems from 20 to 100MW (current 4-6 MW – 12 MW for Kei)

How to build such systems and make them energy sustainable/responsible ?

- Hardware can help (component by component)
- Software must be adapted to be scalable but also more energy efficient
- Usage must be energy aware

How to decrease the energy consumption without impacting the performances?

- How to monitor and to analyze the usage and energy consumption of large scale platforms?

- How to design energy aware software frameworks ?

- How to help users (admins, application and system designers, end users) to express theirs Green concerns and tradeoffs between performance and energy efficiency ?

- **Methodology** : Proposing a generic model able to be derivated onto different scenario (Grids, Clouds, Networks) - Designing software solutions for infrastructures - Simulating and Validating at medium and large scale

Context : Reservation-based systems

Every usage is based on a reservation (resources, duration, deadline):

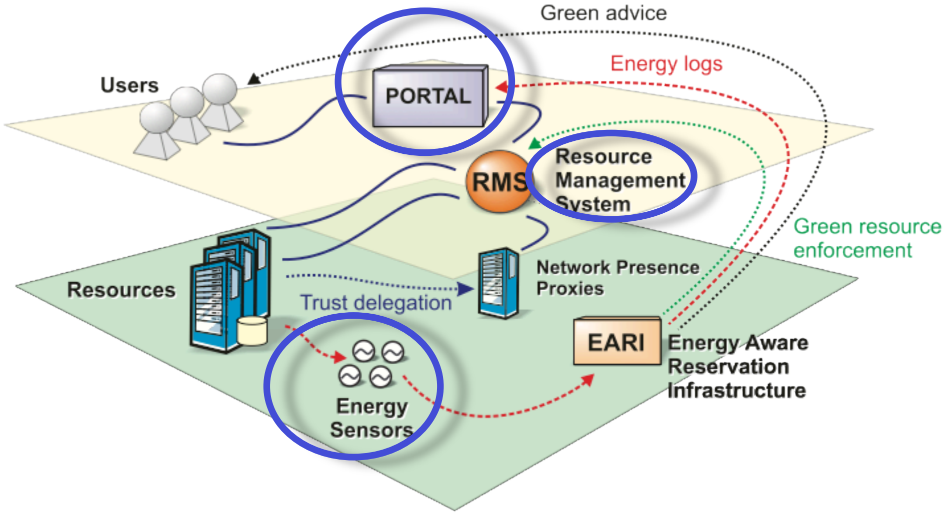
- Reserving cores in exascale (HPC) and Grids
- Reserving Virtual machines time in Clouds
- Reserving Bandwidth in large transport of data

Leverages:

- Finding and powering the optimal number of resources in front of needs of applications
  - Exascale and Grids : switching on/off physical components
  - Clouds : freezing and migrating Vms
  - Networks : lighting or switching off paths, nics, links, routers, LPI
- Adapting « speed » (and consumption) to the need of applications/users
  - HPC, Grids : dvfs
  - Clouds : tuning, capping
  - Networks : adaptive link rate

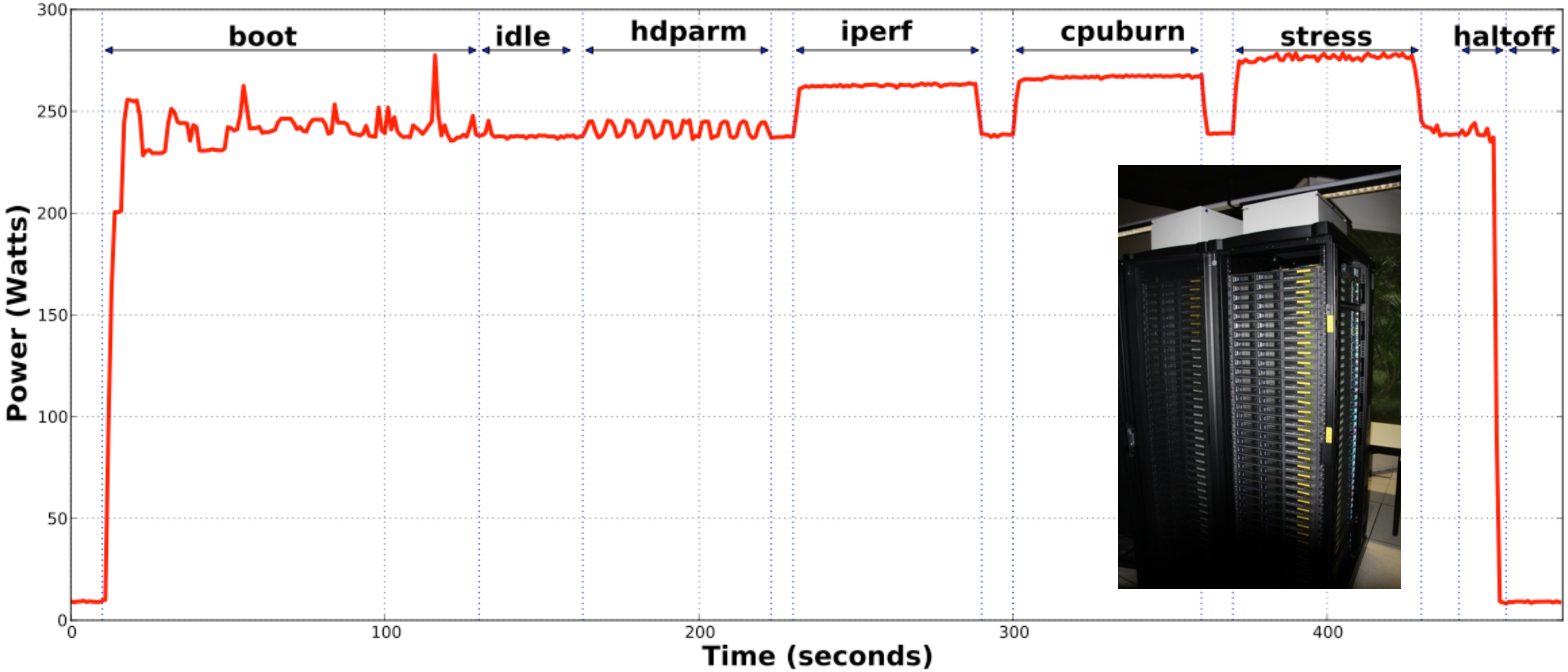
The ERIDIS model : Energy-efficient Reservation Infrastructure for large-scale DIstributed Systems

- Collecting and exposing : usage, energy profiling of applications and infrastructures
- Predicting usage of infrastructures
- Expressing and Proposing : to deal with tradeoffs between perf and energy, Green Policies
- Agregating resources reservations and usage in time and space
- Enforcing Green leverages : switch on/off or adapt performances



Application to Clouds and Networks : Hermes : **High-level Energy-awaRe Model for bandwidth reservation in End-to-end networkS (Globecom2011)**

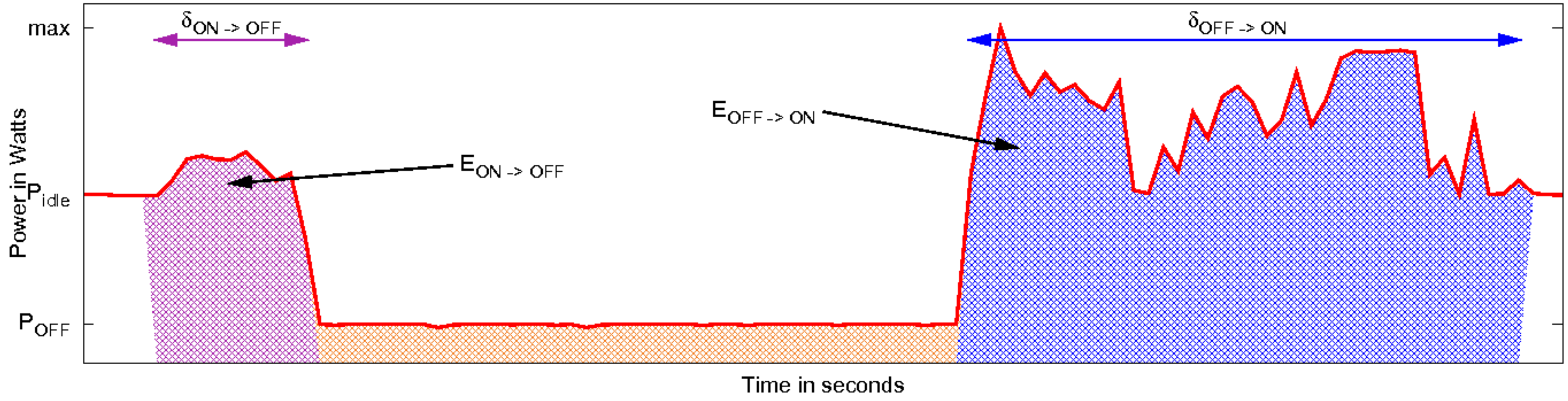
Profiling the energy consumption of applications



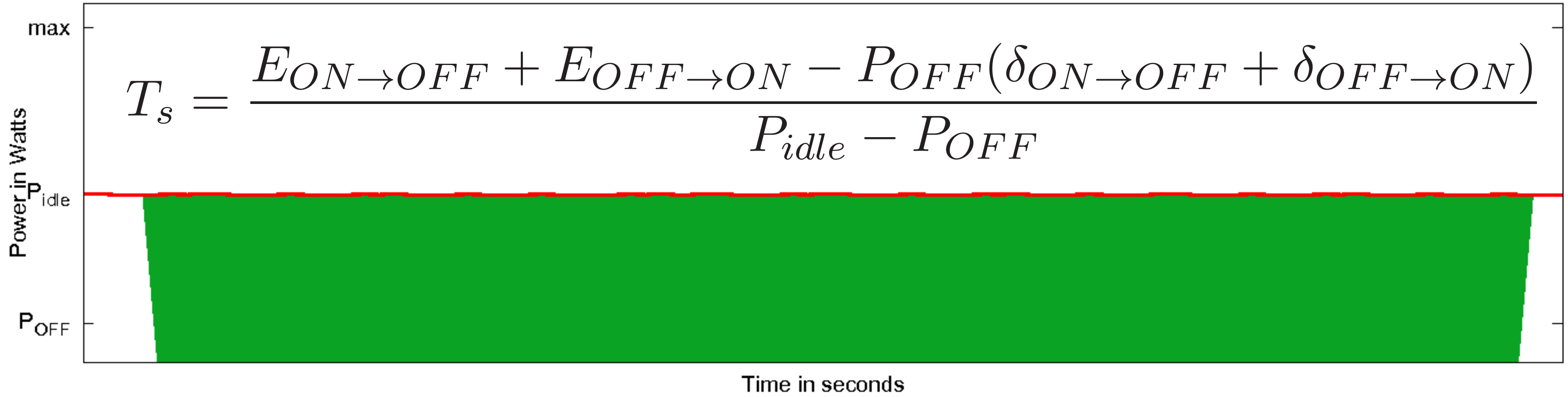
Idle consumption of a node can account up to 90% of the power it consumes when performing a task!

On/Off model

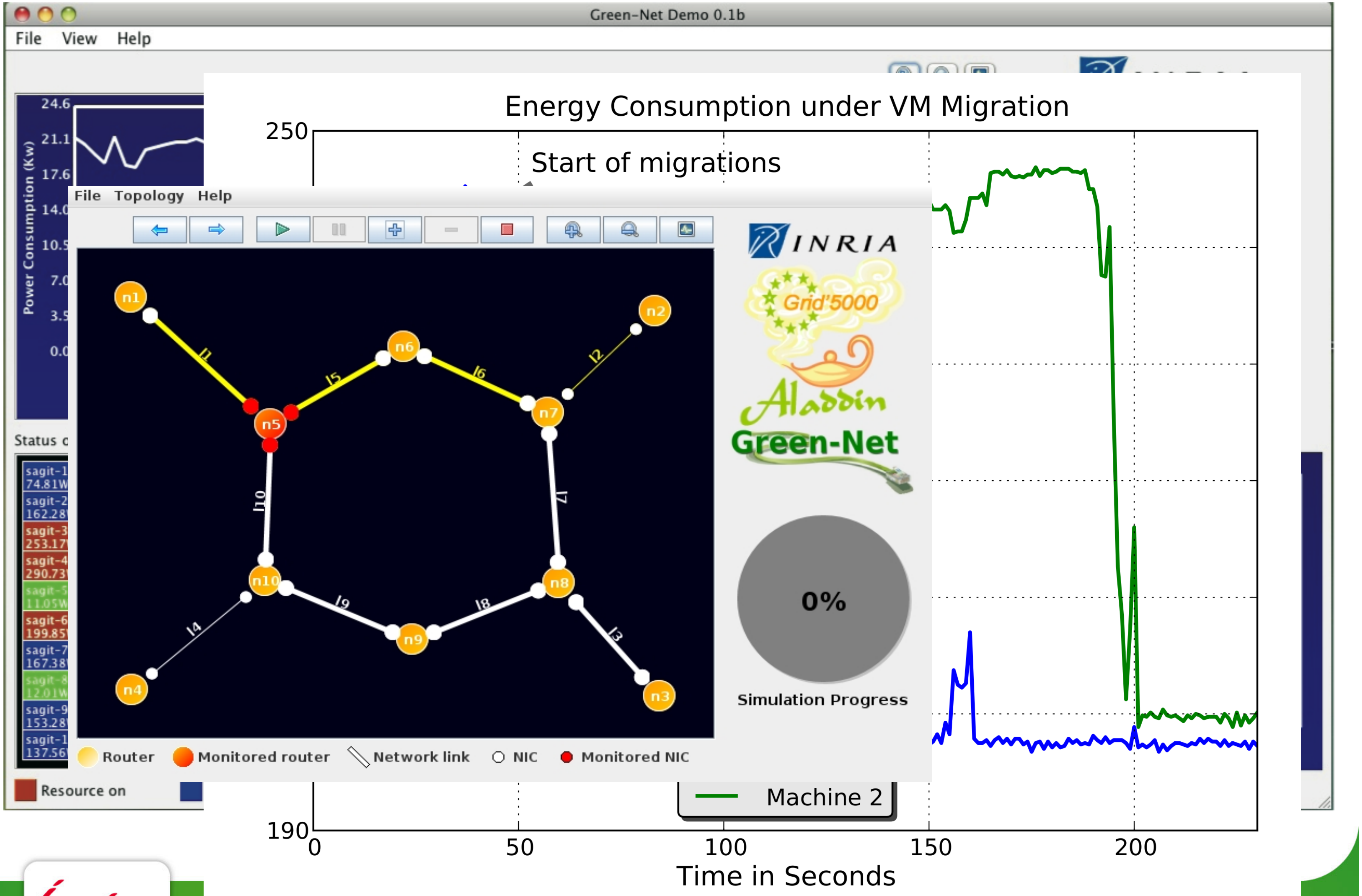
Consumption of the resource if it is switched off and on



Consumption of the resource if it stays idle



$$T_s = \frac{E_{ON \rightarrow OFF} + E_{OFF \rightarrow ON} - P_{OFF}(\delta_{ON \rightarrow OFF} + \delta_{OFF \rightarrow ON})}{P_{idle} - P_{OFF}}$$



Green-Net Demo 0.1b
File View Help
Power Consumption (Kw)
24.6
21.1
17.6
14.0
10.5
7.0
3.5
0.0
Status
sagit-1 74.81W
sagit-2 162.28
sagit-3 253.17
sagit-4 290.73
sagit-5 11.05W
sagit-6 199.85
sagit-7 167.38
sagit-8 12.01W
sagit-9 153.28
sagit-1 137.56
Resource on
Energy Consumption under VM Migration
Start of migrations
250
190
0
50
100
150
200
Time in Seconds
Machine 2
File Topology Help
n1
n2
n3
n4
n5
n6
n7
n8
n9
n10
l1
l2
l3
l4
l5
l6
l7
l8
l9
l10
INRIA
Grid'5000
Aladdin
Green-Net
0%
Simulation Progress
Router
Monitored router
Network link
NIC
Monitored NIC

On-going and future works

Virtualizing Home Gateways at Large Scale (joint work with Bell Labs in GreenTouch)

- HomeGateway to VHGW : 10W per box to <1W per box
- Deporting home gateway services and deploy them in cloud
- Analyzing the Impact on QoS and on energy reduction

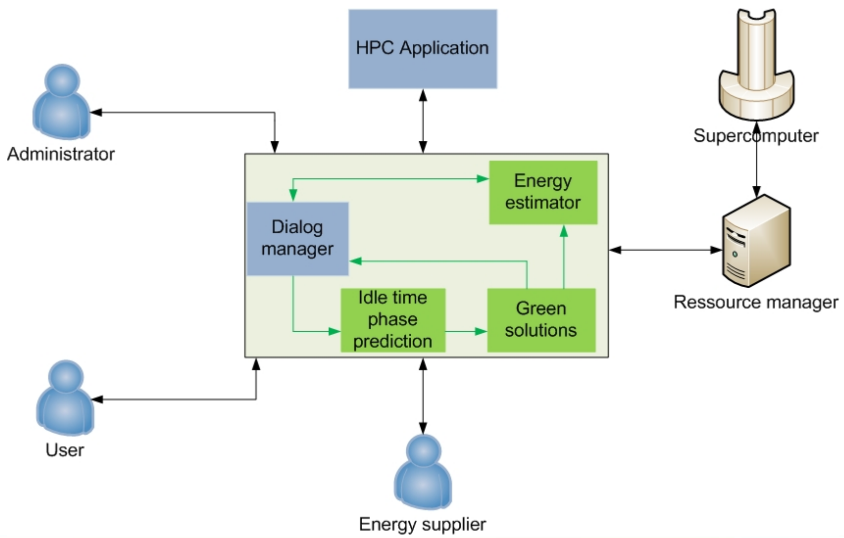
HPC Application Energy profiling (joint work with IRIT, Univ. Toulouse under INRIA Hemera) :

- Large scale applications are difficult to understand (ressource usage)
- Applying green local leverages are not enough for energy reduction
- Profiling for predicting usage of ressources and applying coordinated leverages at large scale – Analyzis of internal ressources usage to derivate the energy consumption

On-going and future works

Green Exascale services (joint work with JLPC - NCSA INRIA and Univ. Urbana Champaign):

- Exaflops machines will embed 100 M of cores with a large agregated energy consumption
- A set of mandatory services will be proposed and supported : fault tolerance, monitoring, barrier, data diffusion...
- Analyzing the energetical impact of such services on applications and infrastructures.
- Building Energy aware version of exascale services (estimator, prediction and leverage modules)
- Designing dialog software framework between actors



Future : Energy efficiency in INRIA Avalon

AVALON : Algorithms and Software Architectures for Service Oriented Platforms

Transverse axis: Energy efficiency in large distributed systems

Keeping the RESO expertise

- Profile and predict energy usage of applications
- Investigate energy usage of (yet to come) virtualized/high density and high performance infrastructures : Clouds, exascale, next generation networks
- Promote and inject energy usage in measurements decision systems
- Apply adapted green leverages at multiple layers (network, cpu)

INVENTORS FOR THE DIGITAL WORLD

Grenoble – Rhône-Alpes

Algorithms and Software Architectures for Service Oriented Platforms

Evolution of computing platforms

- Platforms based on the aggregation of large clusters (Grids), huge datacenters (Clouds), collections of PC (Desktop grids), and/or supercomputers (HPC)
- Different characteristics: performance, energy, size, cost, reliability, quality of service, etc.
- Common challenges: large scale, heterogeneity, volatility, on-demand

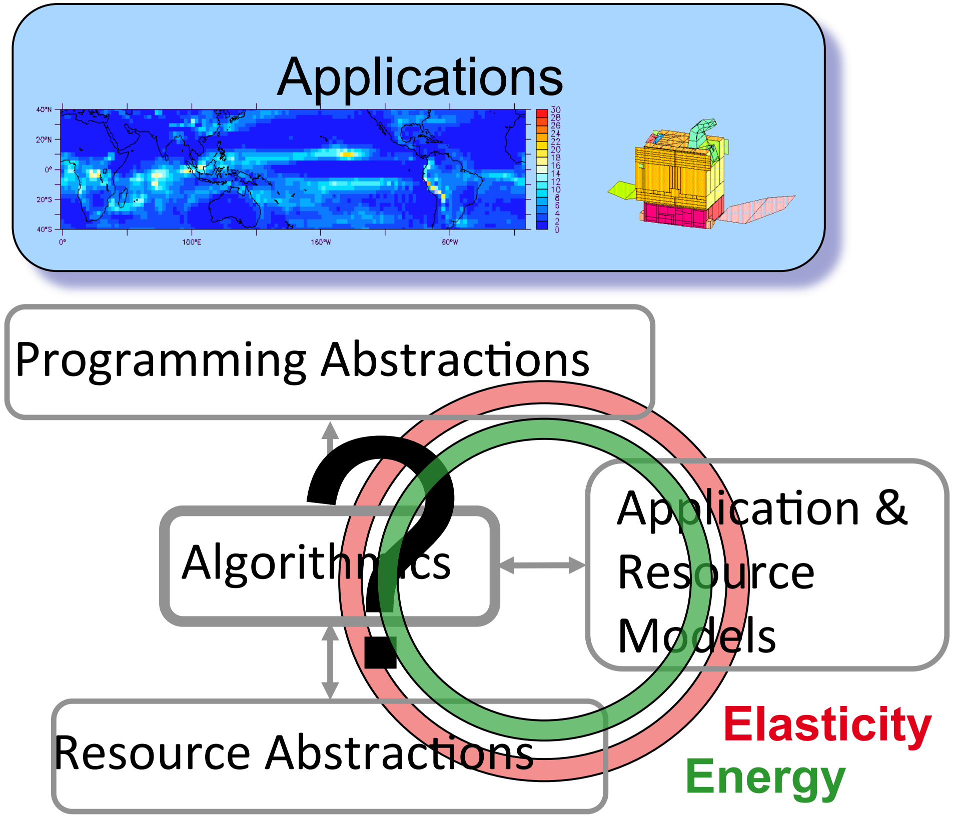
Overall idea

- *Consider the whole system, ranging from resources to applications, as a set of services to be composed*
  - Access transparently to these platforms driven through the use of different services providing mandatory features such as resource discovery, deployment, load-balancing, etc.

Long term goals of Avalon

- Contribute to the design of programming models supporting numerous architecture kinds
- Provide simple to use abstractions of resources
- Efficiently make use of them by mastering the various algorithmic issues involved
- Contribute to the development of middleware frameworks at different levels
- By studying the impact on application-level algorithms

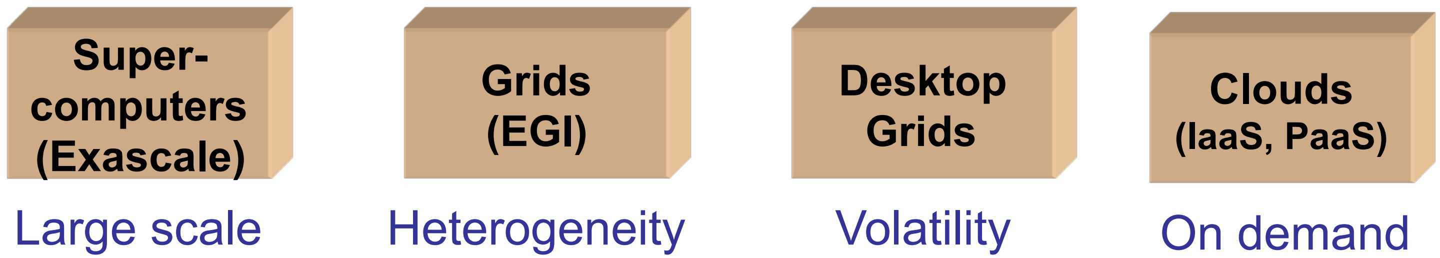
Avalon: Research Activities



CPU/data-intensive Scientific Applications
- From “simple” to code coupling
  - Complexity
  - Code and language heterogeneity
  - “New” forms of interactions (MR)

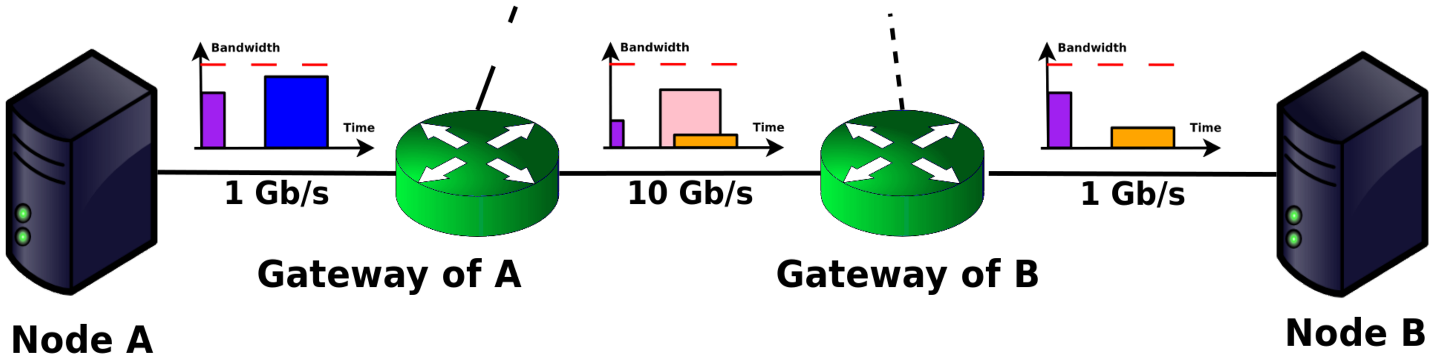
Objectives
- Expressiveness simplicity
- Application portability
- Resource specific optimizations
  - Elastic resource management
  - Energy consumption

Team strength
- Complementary kinds of expertise



HERMES : High-level Energy-awaRe Model for bandwidth reservation in End-to-end networkS

- Switching off unused parts of the network : NIC, routers, links
- Distributed network management
- Energy-efficient scheduling with reservation aggregation
- Usage prediction to avoid on/off cycles
- Minimization of the management messages
- Usage of DTN (Disruptive-Tolerant Network) for network management purpose



Hermes results

- Network simulated: 500 nodes, 2 462 links.
- Random Network (Molloy & Reed method)
- All the nodes can be sources and destinations.
- Time to boot: 30 s.; time to shutdown: 1 s.
- 1 Gbps per port routers

Component	State	Power
Chassis	ON	150 W
	OFF	10 W
Port	1 Gbps	5 W
	100 Mbps	3 W
	idle, 10 Mbps	1 W

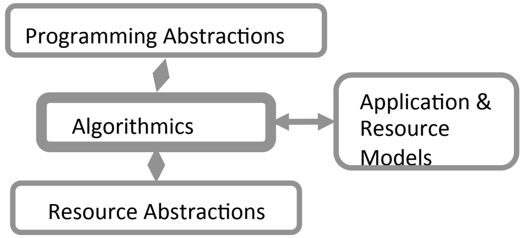
- 31% workload : Energy consumption in Wh

Scheduling	No off	First	First green	Last	Last green	Green
Average	412 306	205 270	203 844	204 949	196 260	203 342
Standard deviation	2 685	2 477	1 938	2 375	2 695	2 145
Accepted volume (Tb)	2 148	2 148	2 128	2 014	1 853	2 149
Cost in Wh per Tb	191.92	95.55	95.78	101.74	105.92	94.60

- Cost in Wh per Tb
- Compared to current case (no-off), HERMES could save 51%, 46% and 43% of the energy consumed depending on the workload

Workload	No off	First	First green	Last	Last green	Green
31%	191.92	95.55	95.78	101.74	105.92	94.60
46%	149.84	81.61	81.95	87.74	92.40	80.63
61%	130.45	74.73	74.91	80.09	84.63	73.79

Avalon: Four Research Axes



Programming abstractions

- Composition based programming models such as component/workflow models.
    - "Transform" resource-aware applications into resource-aware executions
- Domain specific languages

Resource abstractions

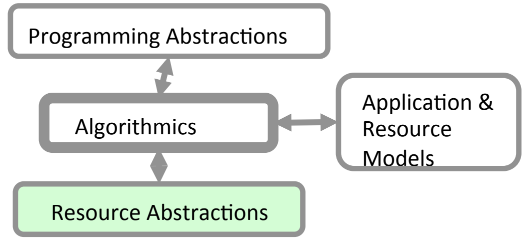
- Hybrid platform management
- Service deployment, service composition and orchestration, service discovery
- Data service and execution of compute/data-intensive applications

Application and resource models

- Accurate/realistic models for application and execution infrastructures
    - Application profiling models and tools
- Energy consumption models

Algorithmics

- Trying to narrow the gap between programming models and resource management systems
- Understand and define the abstraction offered by RMS so that they can achieve their goals (security, fairness, energy, etc.) while enabling advanced programming models to participate to the resource selection



Resource Abstractions (1/2)

Involved Researchers

- **E. Caron**, **G. Fedak**, **L. Lefevre** , **J.-P. Gelas,** O. Glück, C. Perez, F. Suter

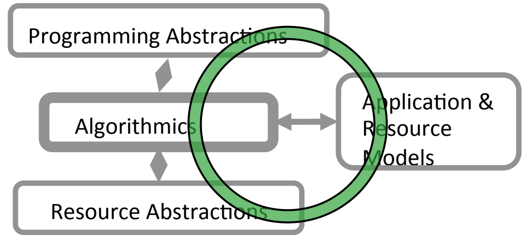
Distributed Computing Infrastructures

- Many different DCIs with many different characteristics
  - Performance, reliability, prices, energy consumption, etc.
- Variety of access usages
  - Batch scheduler, reservation, on-demand, best-effort, virtualized, etc.
- Question: how to provide relevant abstractions to allow efficient resource usage?

Challenges

- Adequate resource management services
  - Large scale, heterogeneous, volatile, elastic
- Combining several DCIs together
- Feedback on how applications make use of resources
  - Energy monitoring

Transverse Topic: Energy Efficiency



Involved Researchers

- E. Caron, G. Fedak, J.P. Gelas, O. Gluck, L. Lefevre, F. Suter

Energy usage and efficiency is not taken into account since the design of systems

- Energy is becoming the major constraint
- Many different DCIs with many different characteristics
- How to express QoS w.r.t performance and energy efficiency?

Challenges

- Profile and predict energy usage of applications
- Investigate energy usage of (yet to come) infrastructures
  - Clouds, exascale, next generation networks
- Promote and inject energy usage in measurements decision systems
- Apply adapted green leverages at multiple layers (network, cpu)