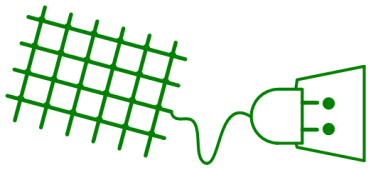


Green-Net

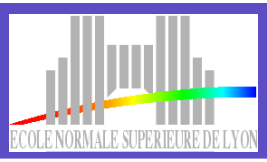
Energy efficiency in large scale distributed systems : the Green-NET initiative

Laurent Lefèvre,

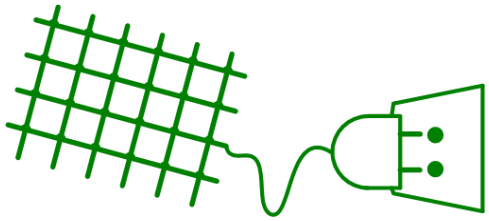
INRIA RESO – University of Lyon, France

laurent.lefevre@inria.fr

OGF25, Catania, 4/3/2009



Green-Net Running project



- **Power aware software frameworks for high performance data transport and computing in large scale distributed systems**
- ARC GREEN-NET : (Action de Recherche Coopérative supported by INRIA)
- Partners teams :
  - IRIT (Toulouse)
  - INRIA MESCAL (Grenoble)
  - INRIA RESO (Lyon)
  - Virginia Tech (USA)

http://www.ens-lyon.fr/LIP/RESO/Projects/GREEN-NET/

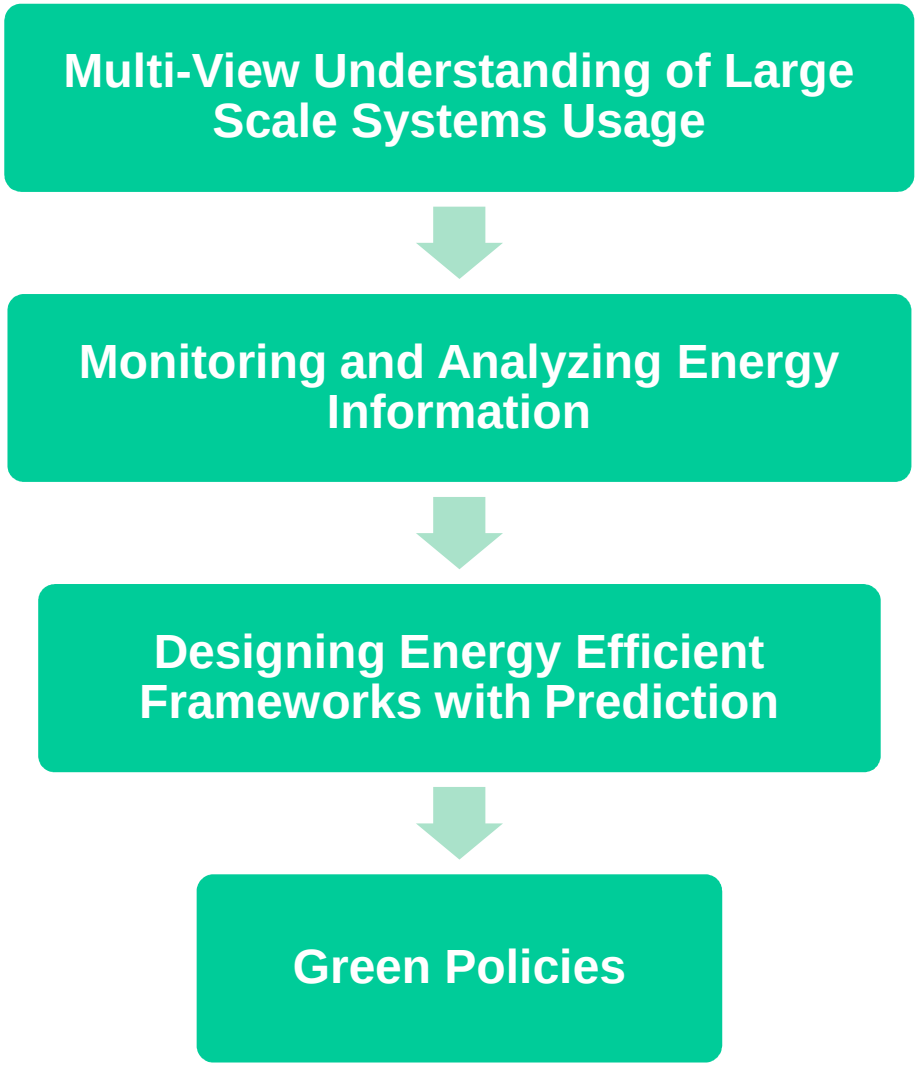
RESO Réseaux haut debit, protocoles et services — VirginiaTech Invent the Future — IRIT Institut de Recherche en Informatique de Toulouse (CNRS INPT UPS UT1) — Grid'5000 — INRIA Rhône-Alpes

Towards Energy Aware Large Scale Systems : open questions

- **How to reduce energy usage without compromising QoE : Quality of Experiment ?**

- How to understand and to analyze the usage of large scale platforms?
- How to apply energy usage models on this experimental usage ?
- How to monitor lively such usage (multiple views (Grids, datacenters, clusters, nodes, services, processes, threads)) ?
- How to design energy aware software frameworks ?

GREEN-NET General Approach



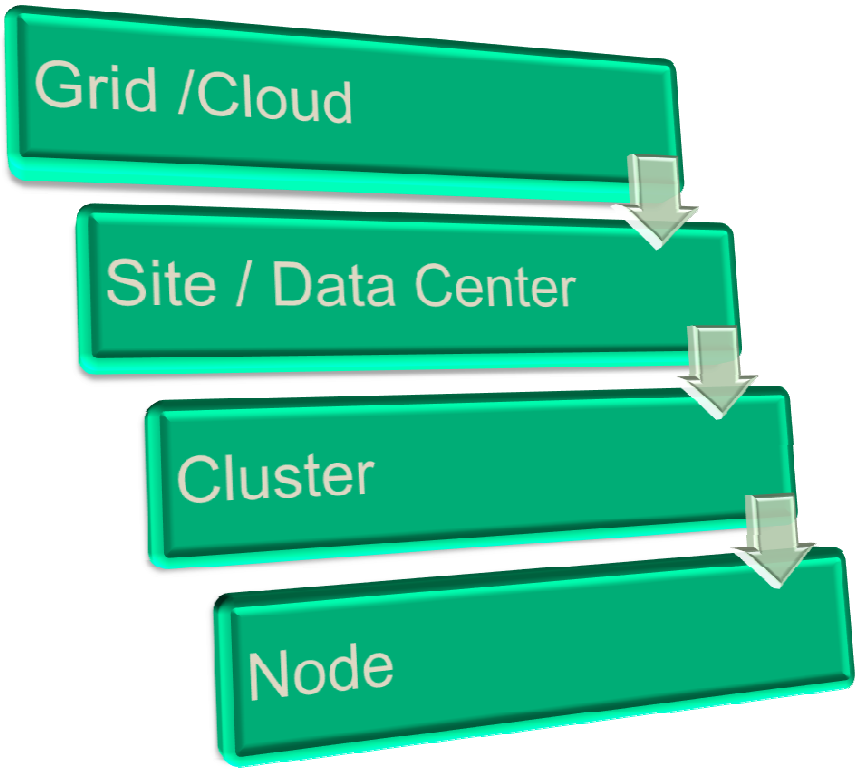
GREEN-NET General Approach

Multi-View Understanding of Large Scale Systems Usage

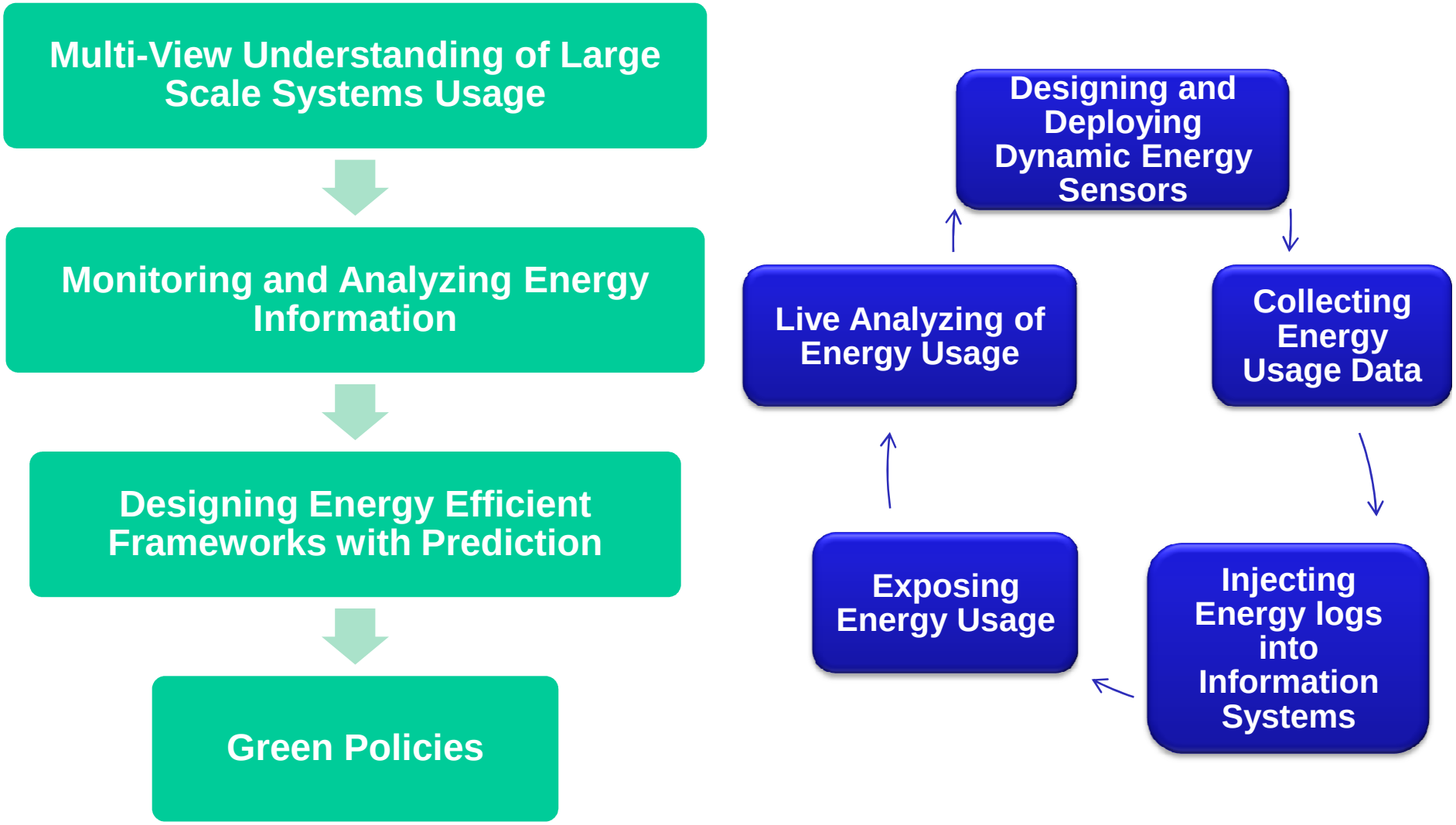
Monitoring and Analyzing Energy Information

Designing Energy Efficient Frameworks with Prediction

Green Policies



GREEN-NET General Approach



GREEN-NET General Approach

Multi-View Understanding of Large Scale Systems Usage

Monitoring and Analyzing Energy Information

Designing Energy Efficient Frameworks with Prediction

Green Policies

Energy Aware Reservation Infrastructure

Prediction Systems

Adapted Schedulers and Resource Management

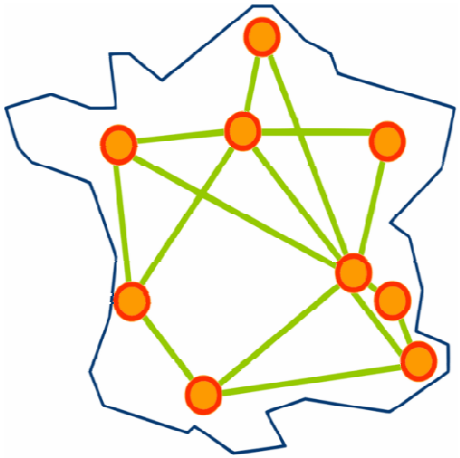
Node and Services Virtualization

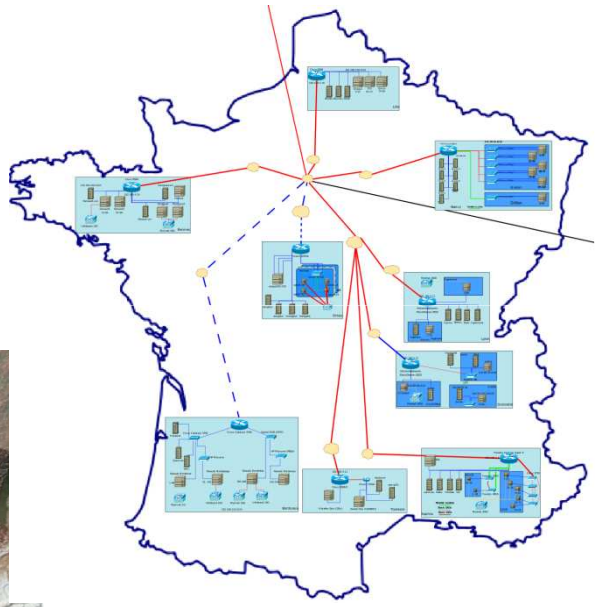
Delegated Trust and Network Presence

Green Policies Support

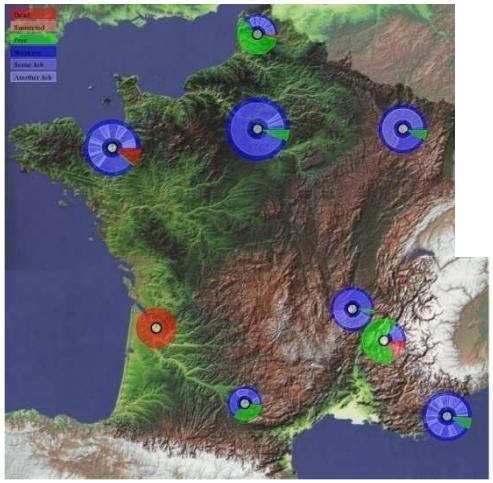
Node Energy Controlers

How an experimental platform is used : the Grid5000 case

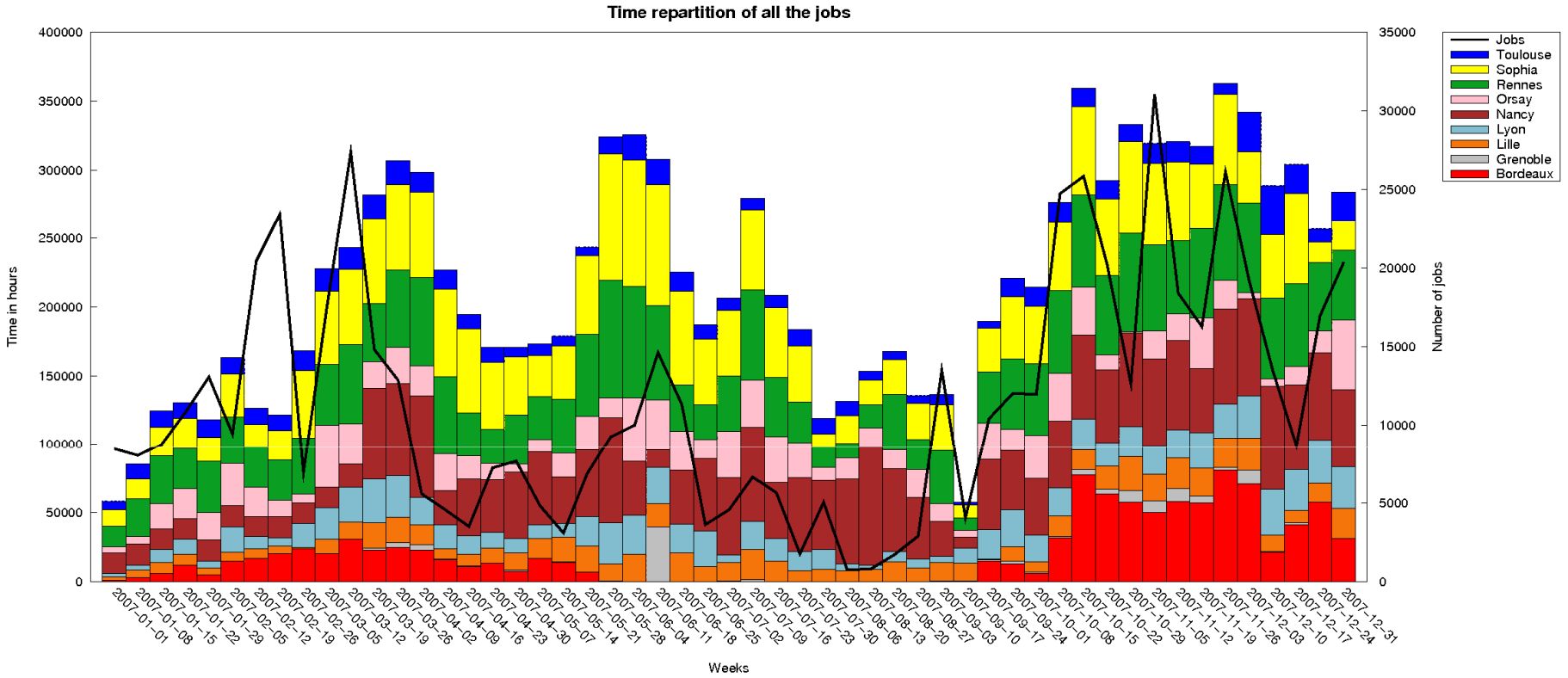




- Experimental testbed for research
- 9 sites geographically distributed in France
- 4000 processors
- Usage : Nodes reservation, image deployment, node reboot, exclusive usage of reserved nodes



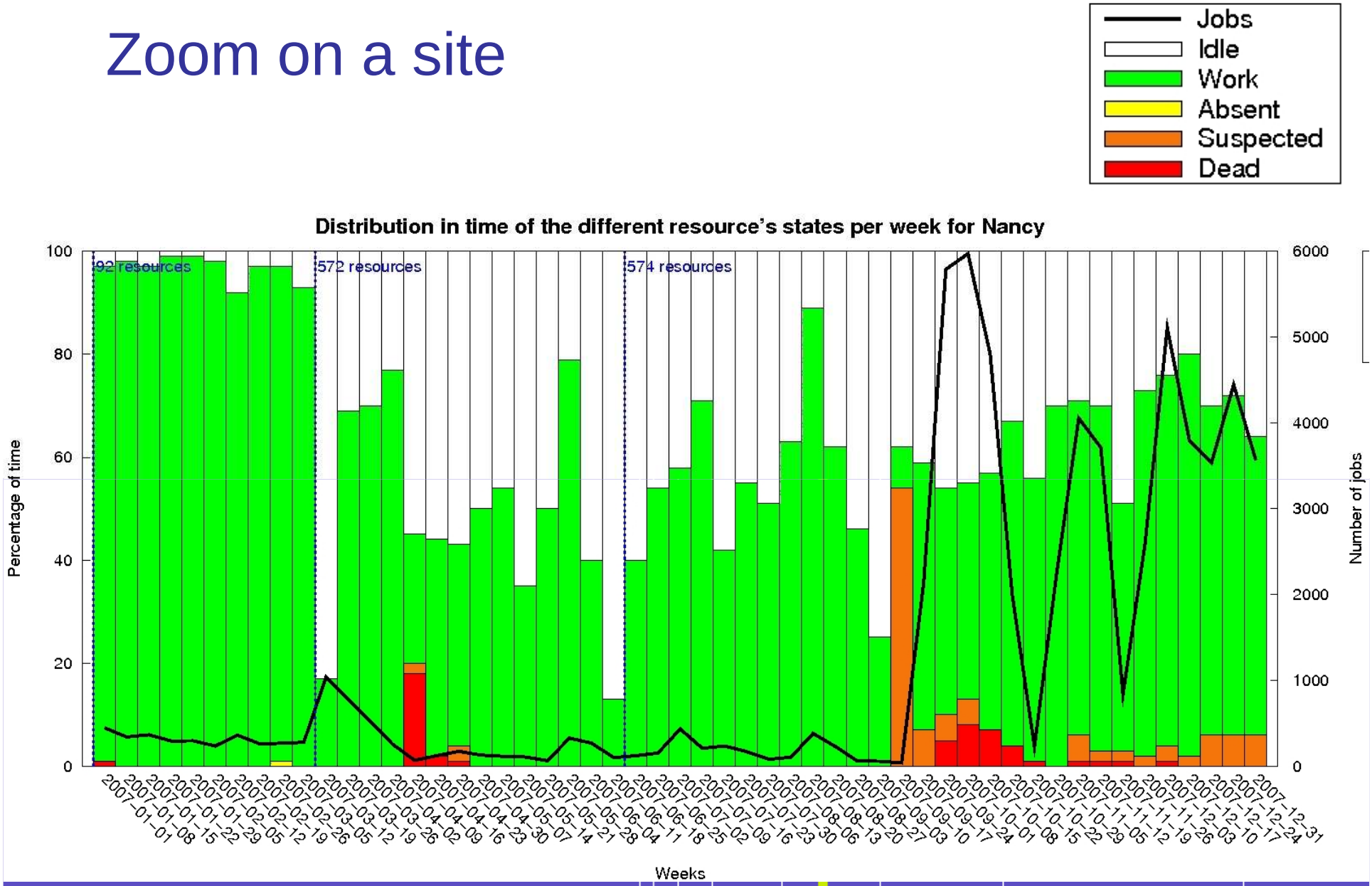
Analysis of global usage



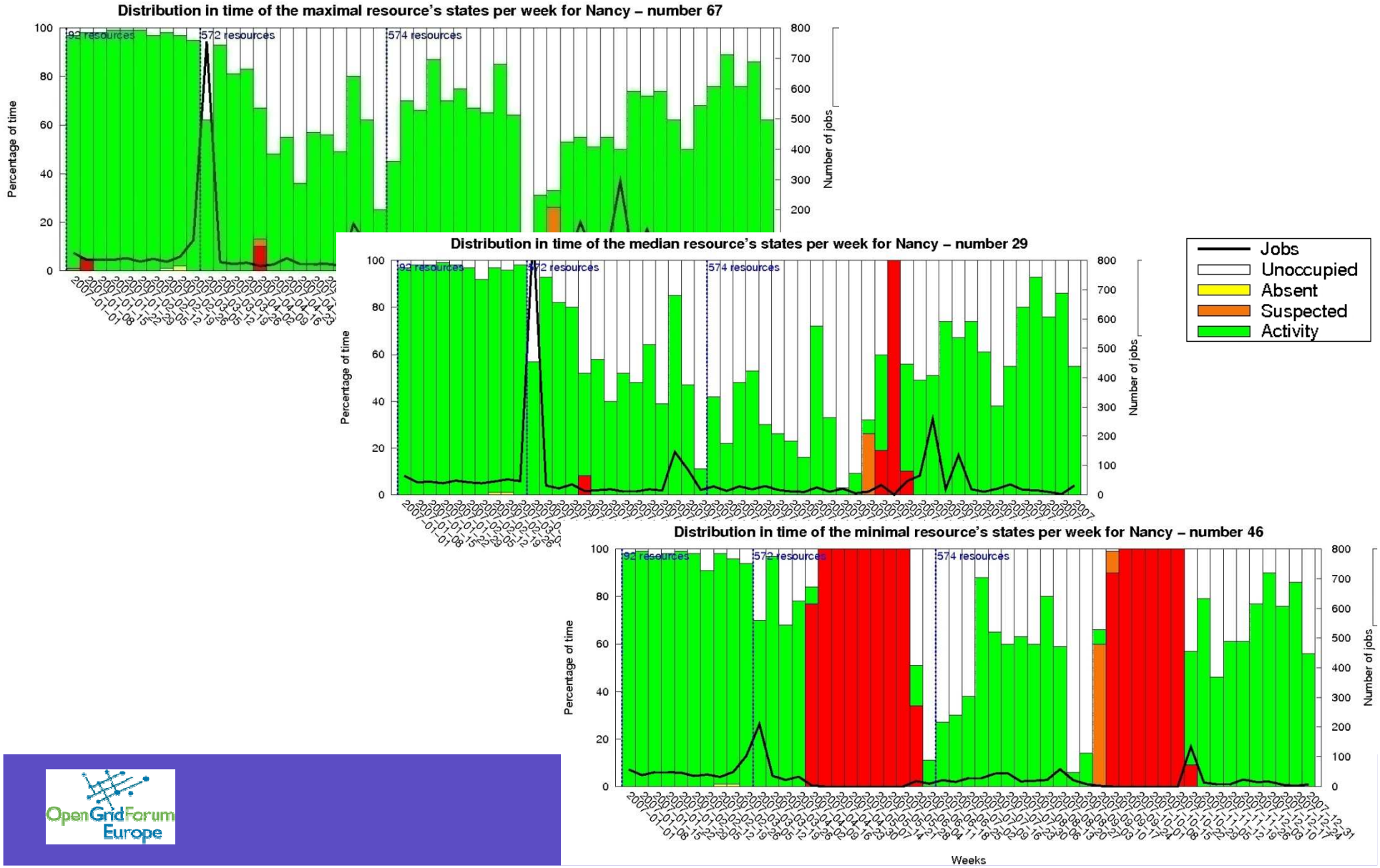
Some low average value but burst support – In average 40% to 70% usage / Operational Grids : 60% to 70% average usage

Global usage is not enough -> need more precise views

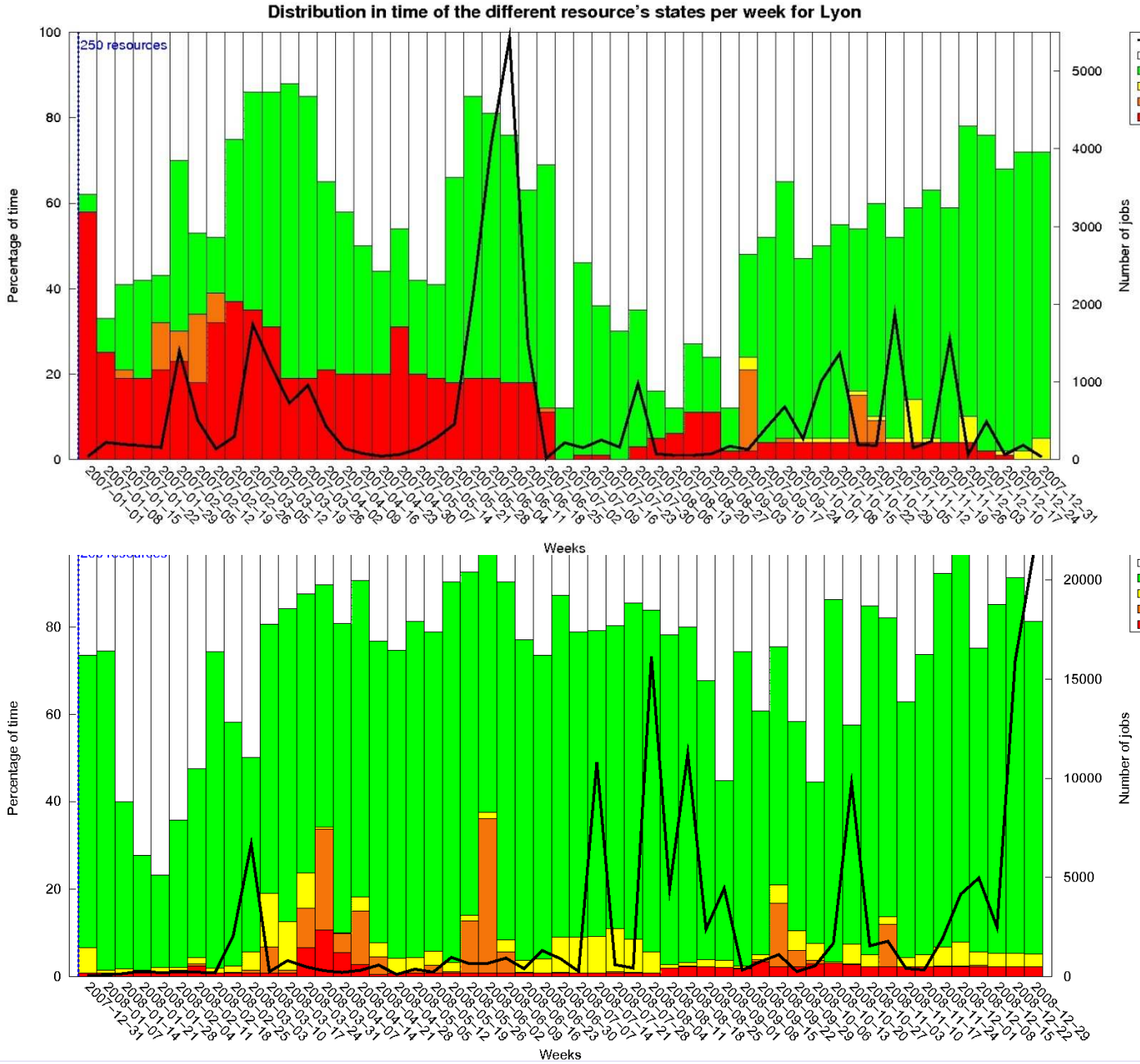
Zoom on a site



Focus on nodes heterogeneity



Platform usage evolution

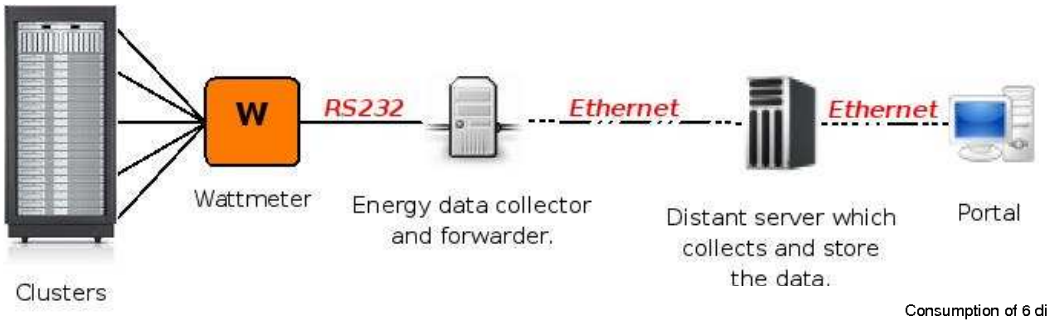


Lessons Learned

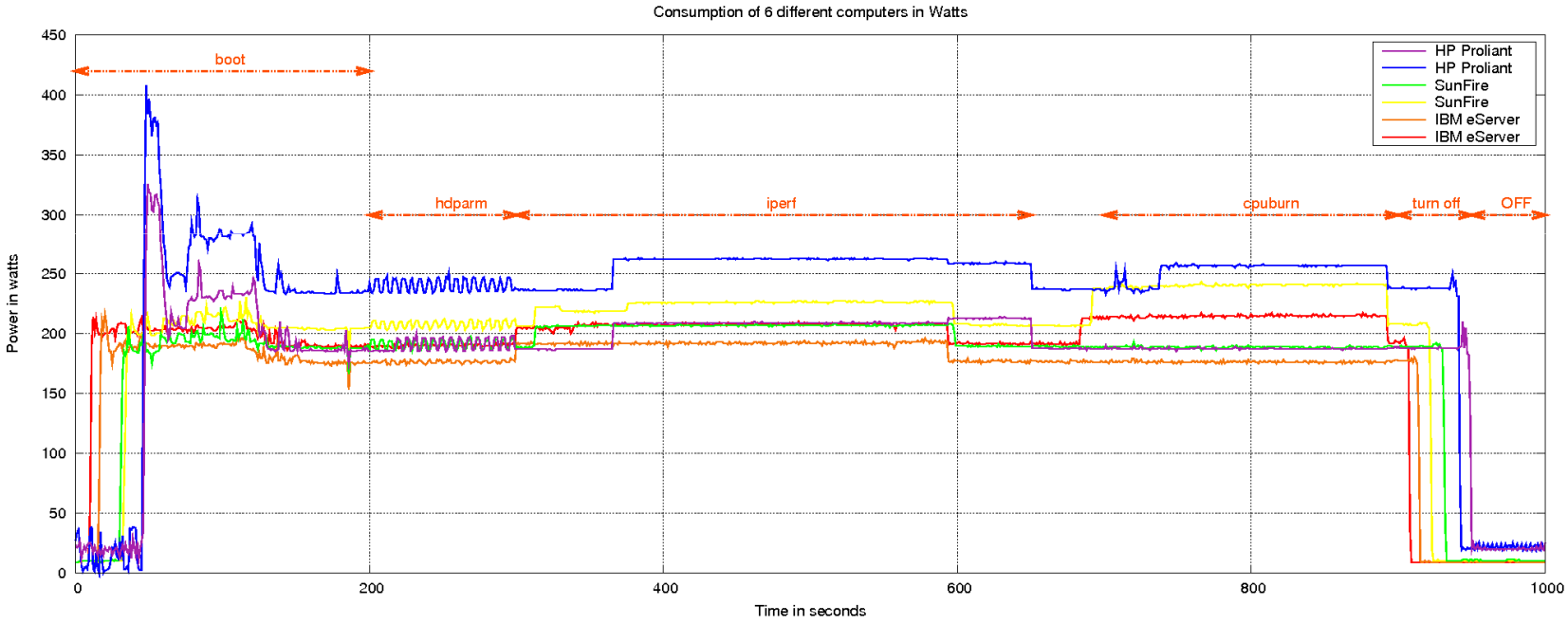
- Significative bursts -> significative gaps!
- A lot of small reservations (experimental platform)
- Significative energy usage while nodes are idle
- Benefit from gaps to propose energy savings models
- No constant HP usage

NRIA RHÔNE-ALPES

Energy usage of nodes



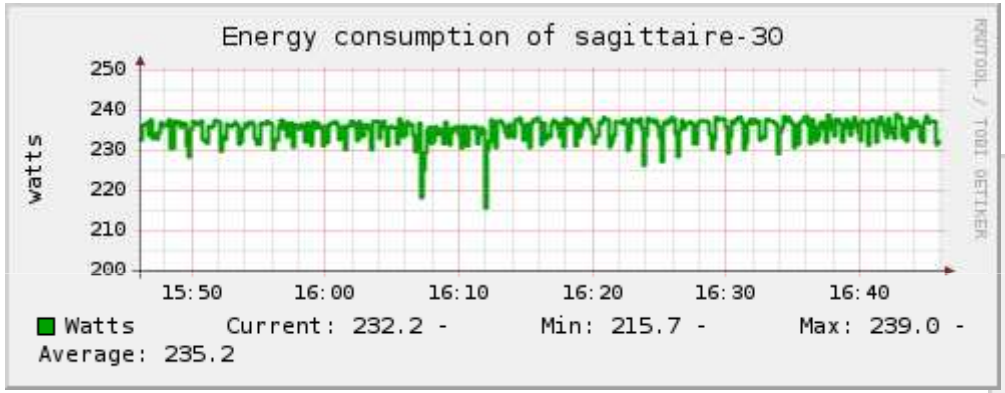
- IBM eServer 325 (2.0GHz, 2 CPU)
- Sun Fire v20z (2.4GHz, 2 CPU)
- HP Proliant 385 G2 (2.2GHz, 2 CPU bicore)

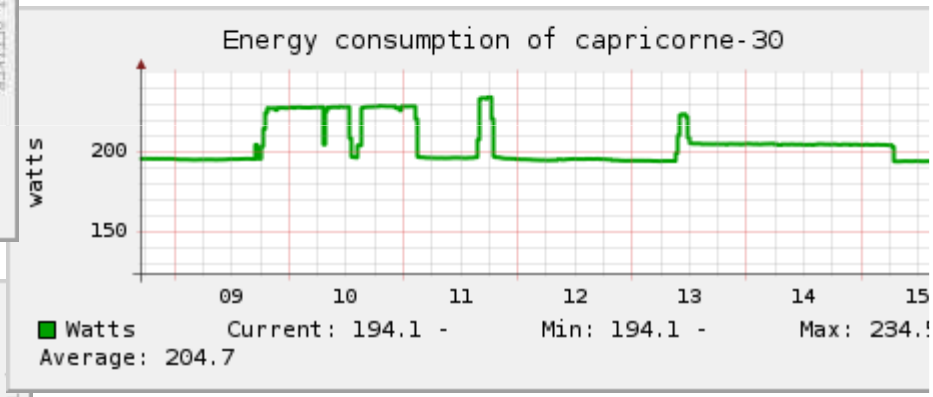


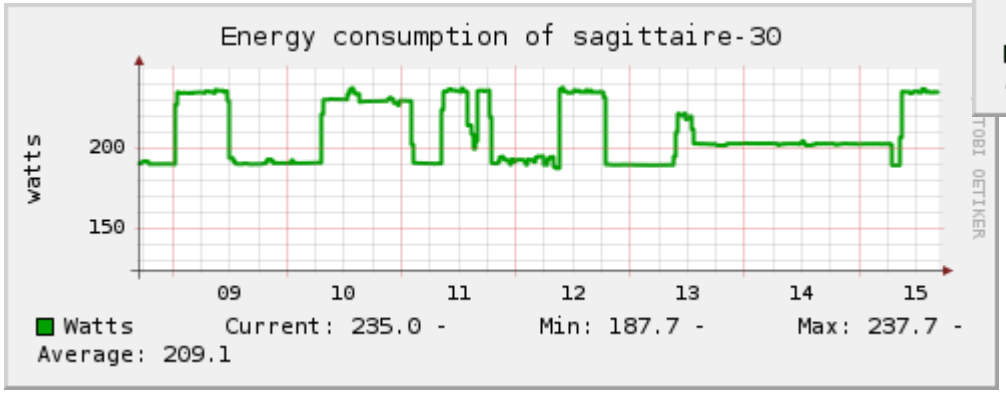
Impact of hardware, usage, location…

Live Energy Monitoring

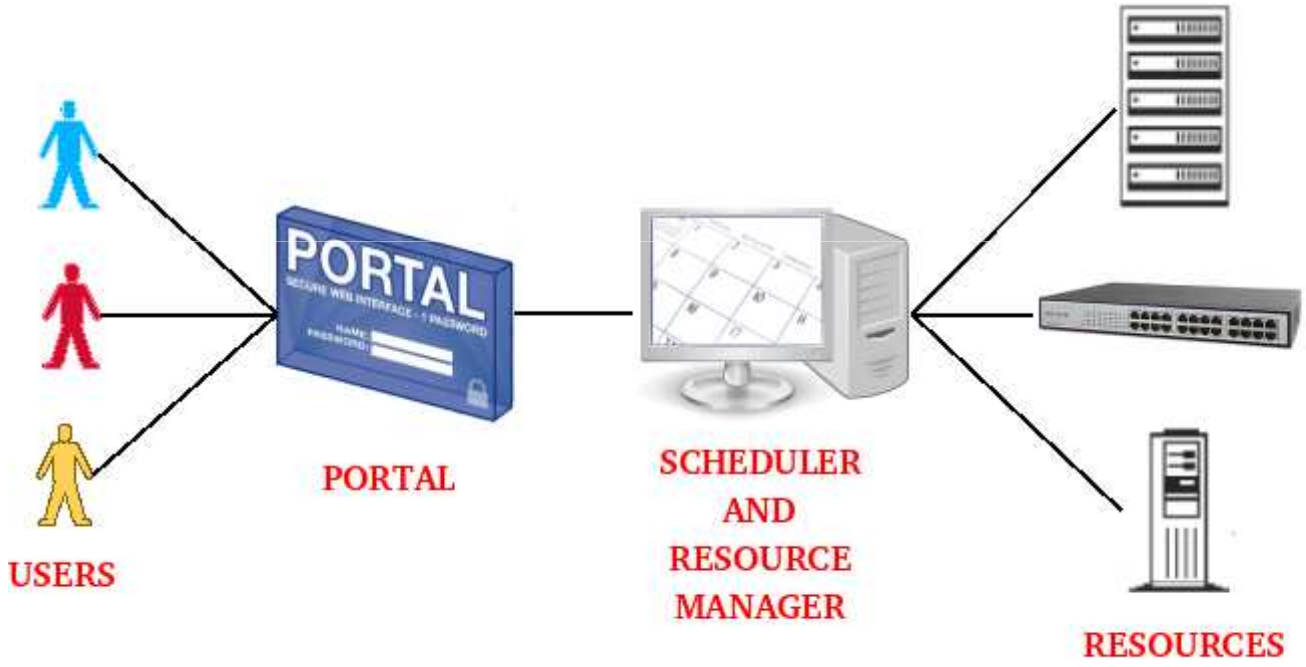
- Feedback for users (days,weeks, month) for the reserved nodes



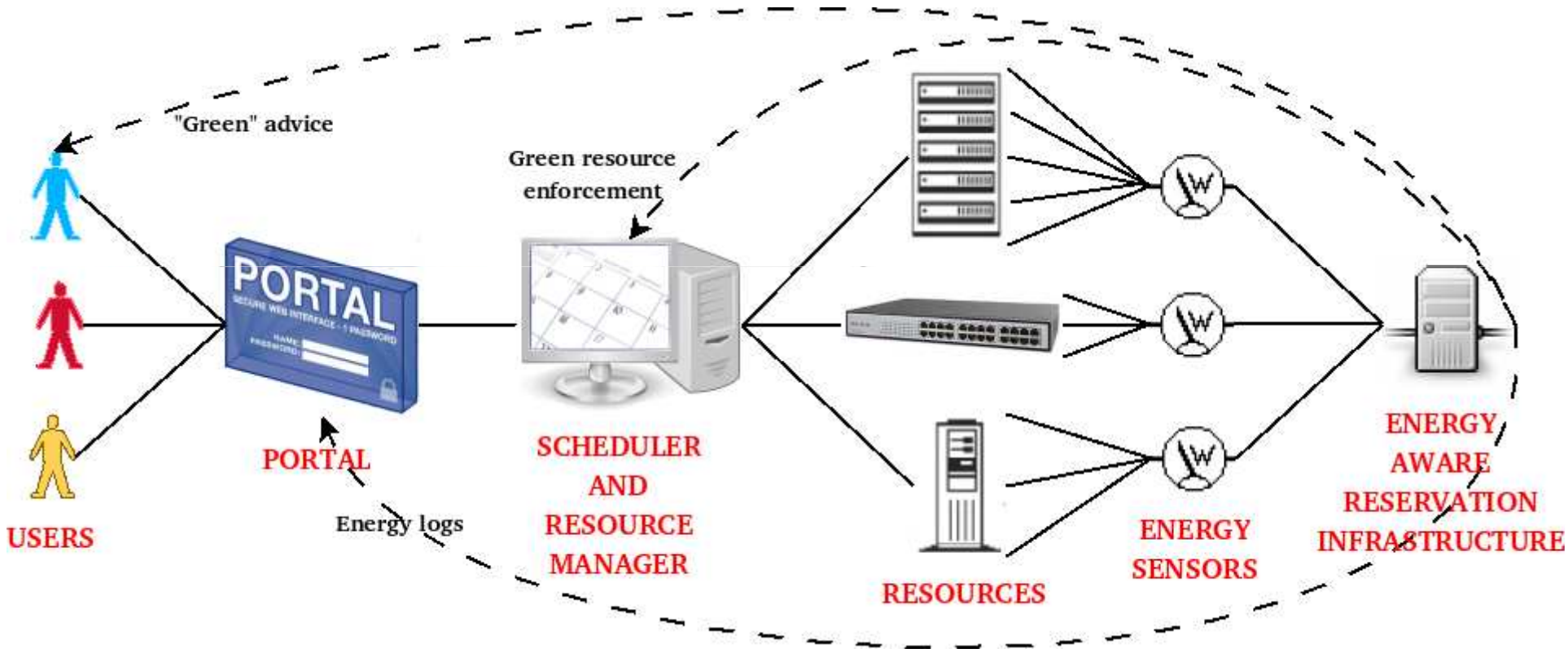




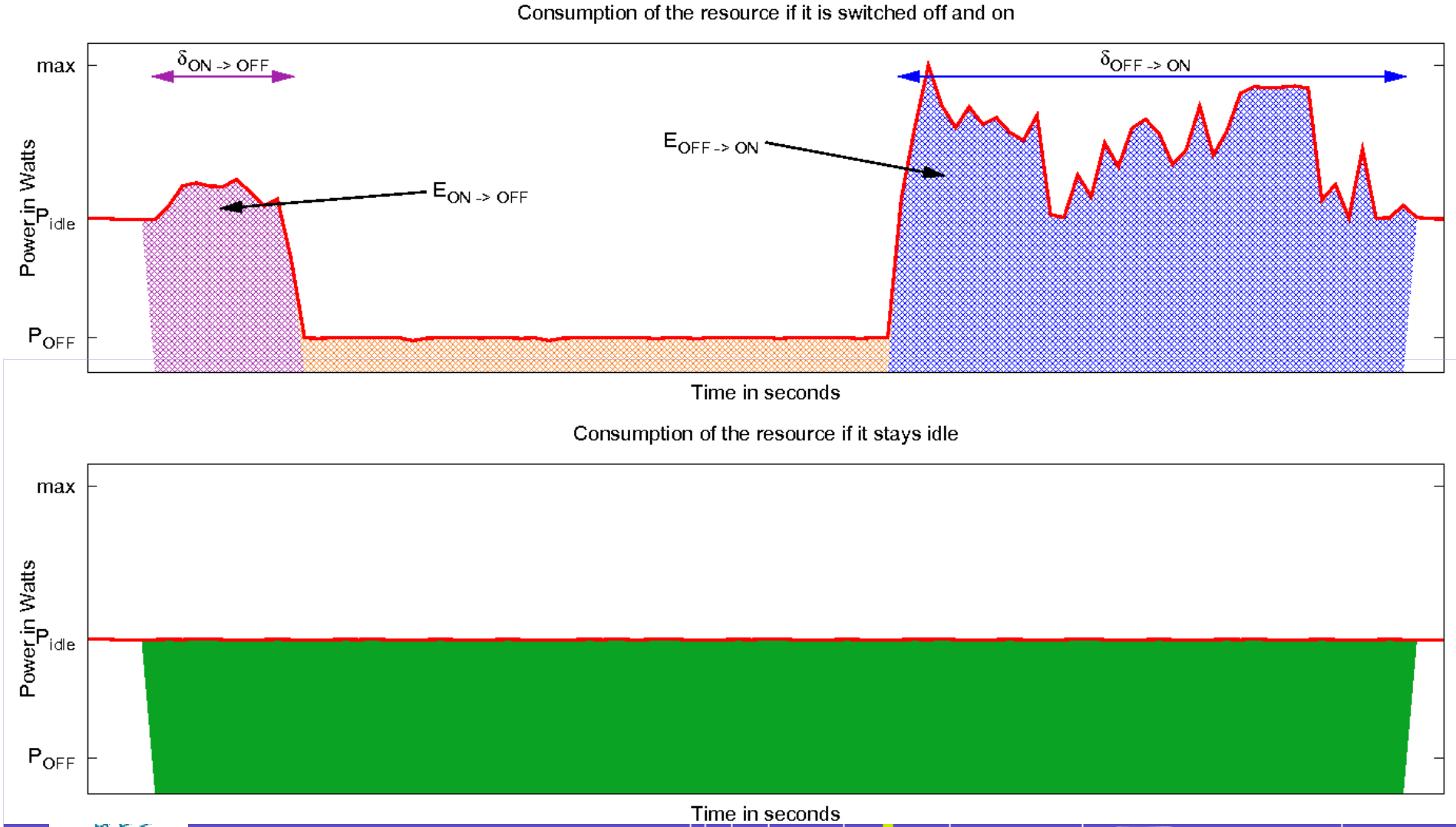
Energy-Aware Reservation Infrastructure



Architecture of an energy aware platform



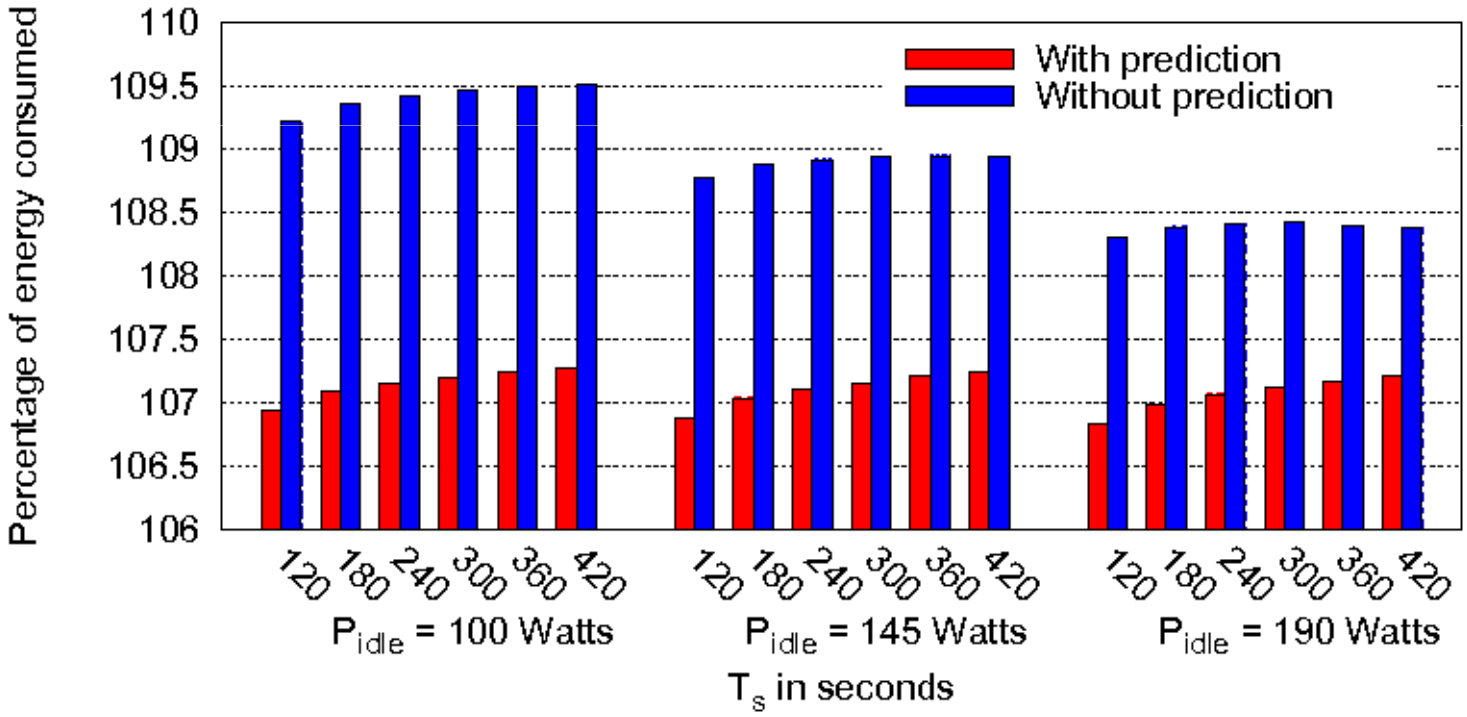
On/Off model? → optimization



On/Off model? → prediction

- At the end of a reservation, we want to know if it is better to shut down the nodes or not.
- Prediction :
  - What :
    - Next reservation (size, duration, start time)
    - Next empty period
    - Energy consumption of a reservation
  - With :
    - Recent history (last reservation) + feedback
    - Recent reservations days + feedback
    - User history + resources

Prediction evaluation based on replay

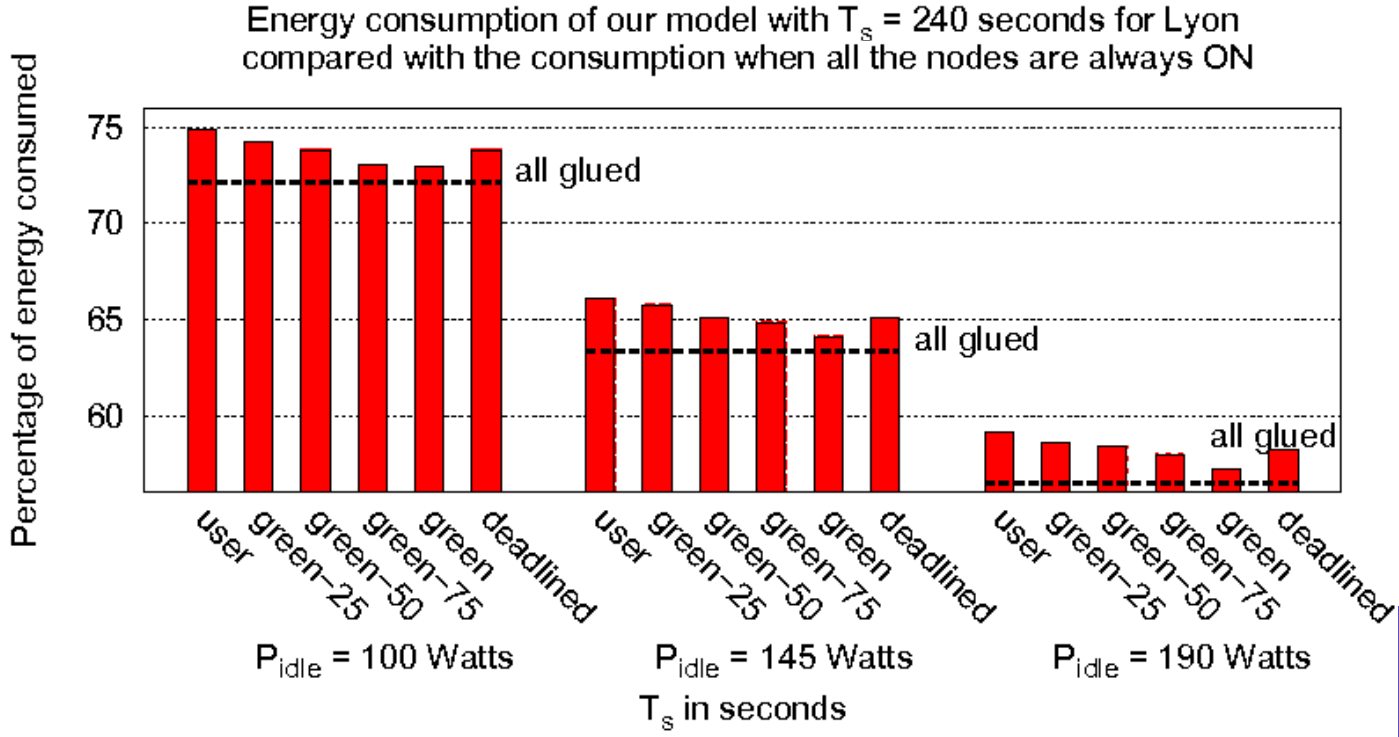
- Example : Bordeaux site (650 cores, 45K reservations, 45% usage)
- 100 % : theoritical case (future perfectly known)
- 

Green policies

- **User** : requested date
- **25% green** : 25% of jobs follow Green advices – the rest follow user request
- **50% green** : 50% of jobs follow Green advices – the rest follow user request
- **75% green** : 75% of jobs follow Green advices – the rest follow user request
- **Fully green** : solution with uses the minimal amount of energy and follow Green advices
- **Deadlined** : fully green for 24h – after :user

Green policies evaluation and future energy savings

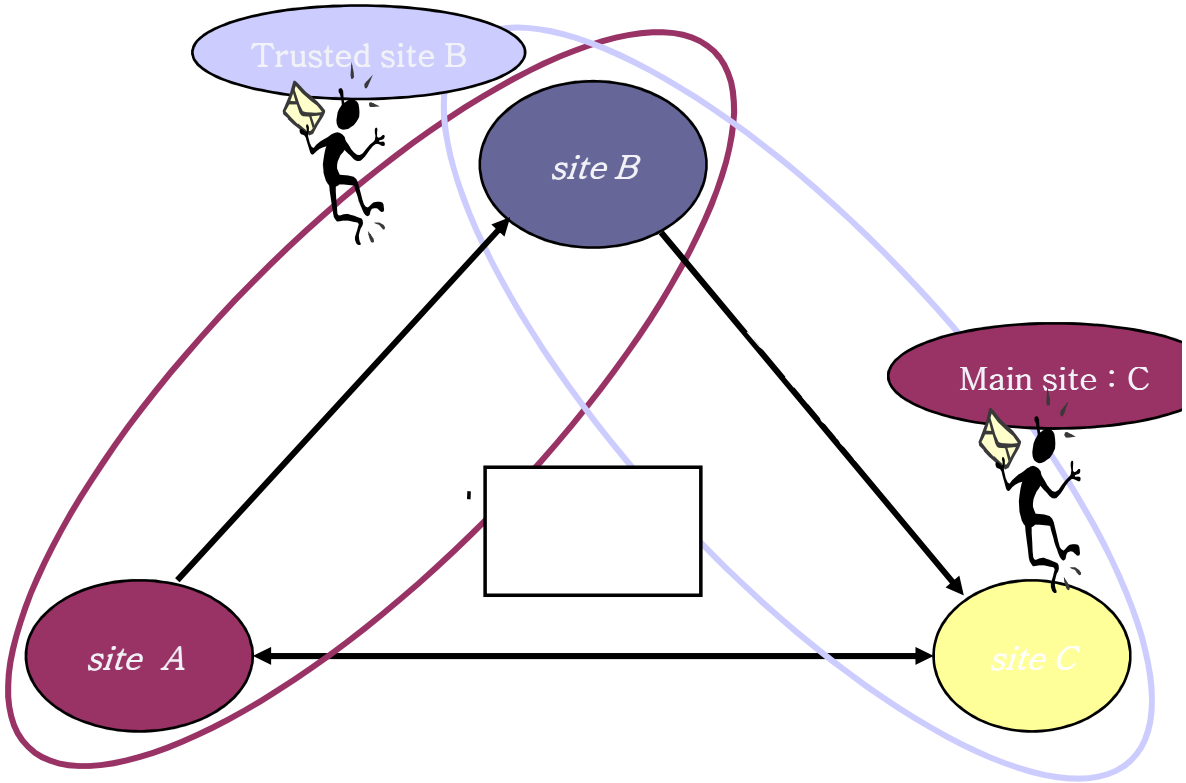
- Example of Lyon site (322 cores, 33K reservations, 46% usage)
- Current situation : always ON nodes(100 %) All glued : unreachable theroritical limit
- For Lyon site : saving of 73 800 kwh for 2007 period
- 1209159 kWh for the full Grid5000 platform (without aircooling and network equipments) on a 12 month periods



Required frameworks for operationnal energy savings

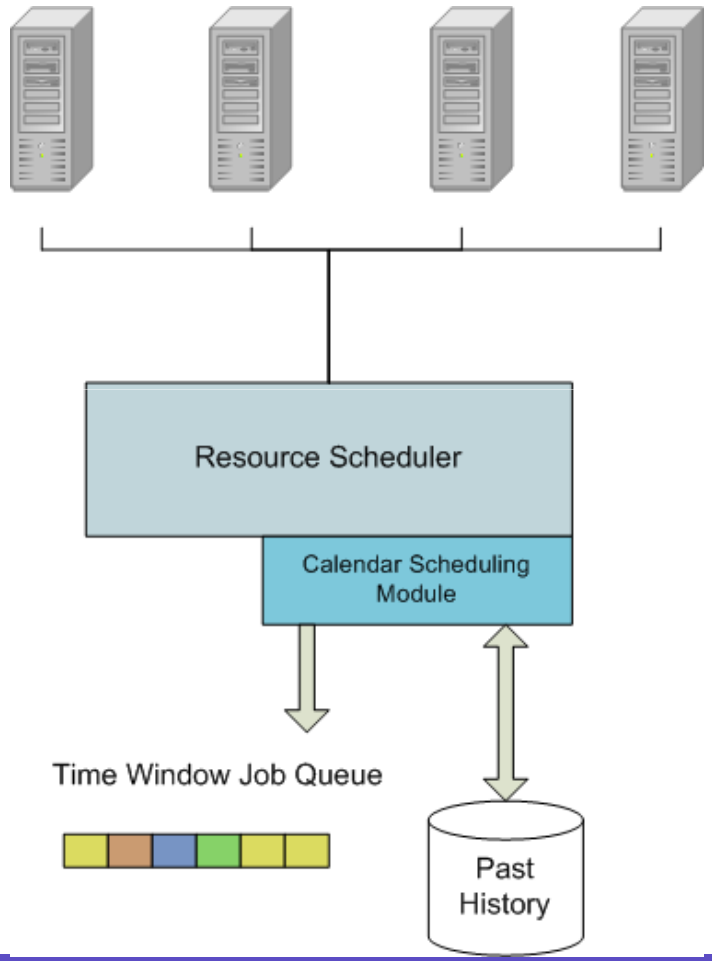
- Supporting network presence for OFF equipments (distributed trust delegations)

- Adapting a Resource Management System for energy efficiency (based on OAR)



Slide provided by IRIT Team (J.M. Pierson, G. Da Costa)

Prediction Based Energy efficient scheduling



Model based upon past, current and future workload of the cluster

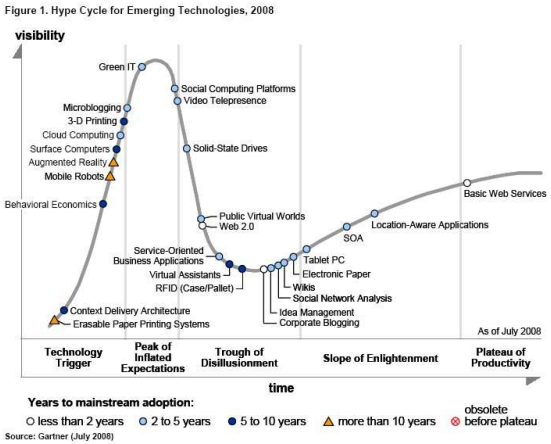
Automated Decisions

- Suspend-To-Ram (for small idle periods)
- Suspend-To-Disk (for medium idle periods)
- Shutdown(for big idle periods)

Users / soft control device power policies during job execution

- Actions like: CPU-GPU Frequency, HardDisk spin down/power off, Network card speed/power off,...
- Used by: Programmed clever applications that provide which device is needed and which can be shut down or function slower

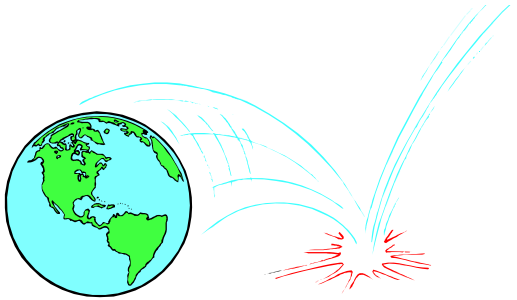
Slide provided by MESCAL INRIA Team (O. Richard, Y. Georgiou)



Conclusions

- How to be energy efficient → how to reduce the watts on a large scale ?
- The energy aspects will /must change the way we design software, protocols and services
- The frameworks and middleware can/must help
- Solutions for large scale / worldwide Grids
- Human factor : are we ready to sacrifice some QoS/performance ?

Current requirements



- Need of energy aware resources and equipments !
- Common practice to access energy logs
- Need of open and adapted energy aware benchmarks
- We are looking for logs ! ☺

Questions ?

laurent.lefevre@inria.fr

http://perso.ens-lyon.fr/laurent.lefevre

http://www.ens-lyon.fr/LIP/RESO/Projects/GREEN-NET/