Hybrid Distributed Computing Infrastructure Experiments in Grid5000: Supporting QoS in Desktop Grids with Cloud Resources

Simon Delamare, Gilles Fedak, Oleg Lodygensky

simon.delamare@inria.fr
INRIA Rhône Alpes
GRAAL team

Grid'5000 Spring School 2011

Plan

Context

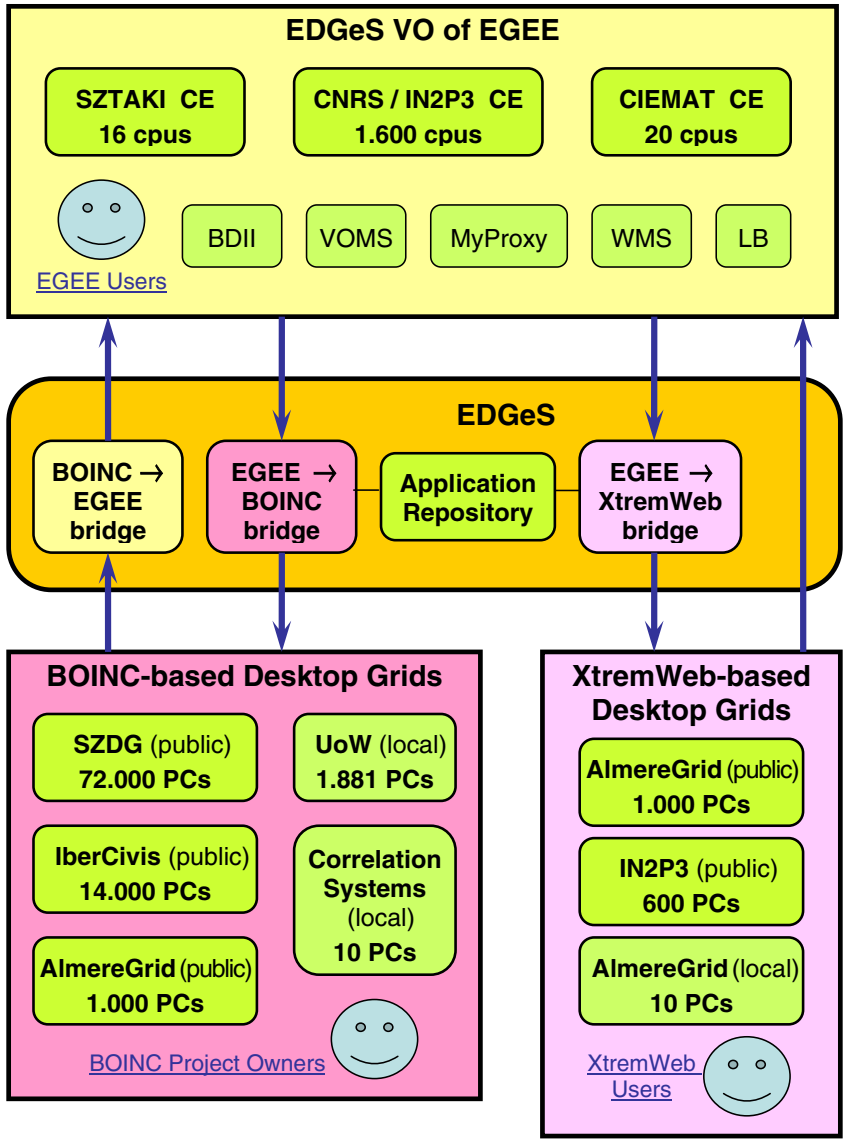
- Growing demand for computing power from scientific communities (large application, large datasets)
- Distributed Computing Infrastructures continues to diversify :
  - Supercomputers, Grids, Desktop Grids, and now Cloud Computing...
  - Different characteristics in term of performance, size, cost, reliability, quality of service etc...

Question : how do we mix them ? According to which criteria/scenario ?

- example : extends Grid infrastructure with Cloud resources to meet a peak demand.
- example : use local Desktop Grid to cut the Cloud resource cost.
- example : use on-demand Cloud resources to improve QoS of Desktop Grid.

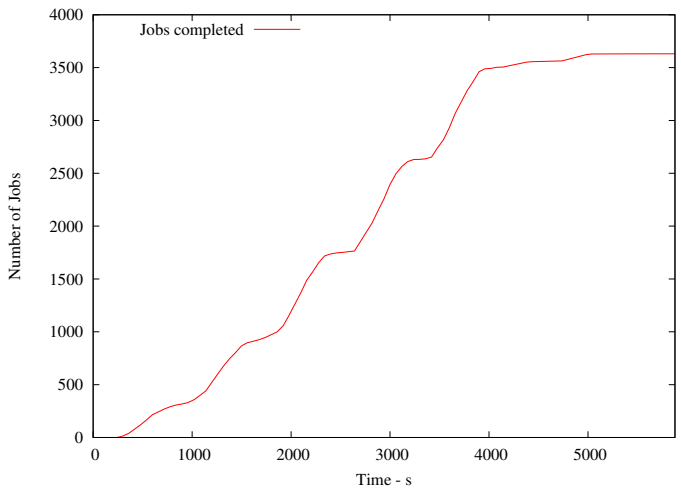
Background : The EU FP7 EDGI

- EDGeS (Enabling Desktop Grid for eScience) : bridges from EGEE to Desktop Grid (BOINC and XtremWeb)
- FP7 EDGI/DEGISCO are new projects to maintain and extend the EDGeS infrastructures
  - new Grids : ARC, Unicore
  - Clouds : Eucalyptus, OpenNebula
- *International Desktop Grid Federation*: 40 partners worldwide.



Motivation for QoS in Best-Effort DCI

- Best Effort DCI: Desktop Grid, Grids' besteffort queues, Amazon EC2 spot instances, ...
- Characteristics: Variable amount of resources, volatility, unpredictability, unannounced departure.
- Low QoS compare to classical DCI → Tail Effect



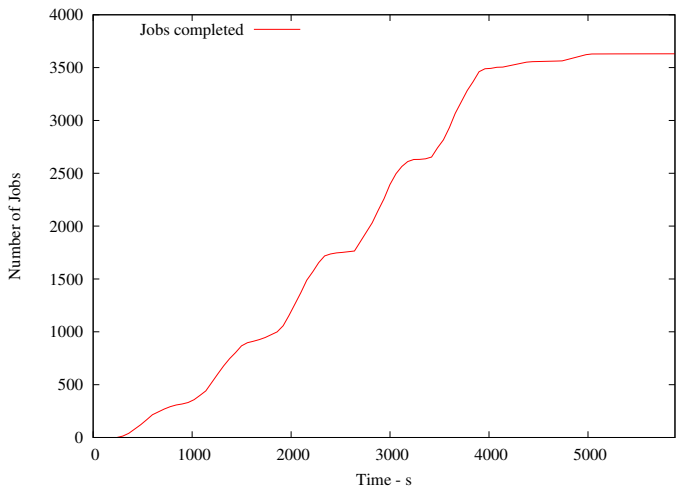
- We define *Quality of Service* as a level of confidence in Bag of Task (BoT) execution :
  - BoT makespan is the time between the BoT is first submitted and the time all the results have been received and validated
  - What can be estimated, predicted, guaranteed ?

Question: how do we provide QoS to users given the dynamism and volatility of the computing resources ?

- Intrinsic approach : improve scheduler for QoS ability
- Extrinsic approach : provide additional dedicated computing resources

Motivation for QoS in Best-Effort DCI

- Best Effort DCI: Desktop Grid, Grids' besteffort queues, Amazon EC2 spot instances, ...
- Characteristics: Variable amount of resources, volatility, unpredictability, unannounced departure.
- Low QoS compare to classical DCI → Tail Effect



- We define *Quality of Service* as a level of confidence in Bag of Task (BoT) execution :
  - BoT makespan is the time between the BoT is first submitted and the time all the results have been received and validated
  - What can be estimated, predicted, guaranteed ?

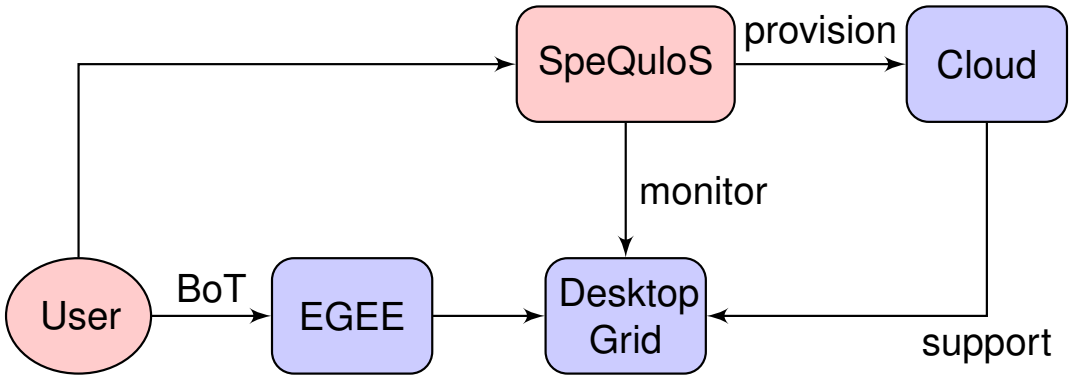
Question: how do we provide QoS to users given the dynamism and volatility of the computing resources ?

- Intrinsic approach : improve scheduler for QoS ability
- Extrinsic approach : provide additional dedicated computing resources

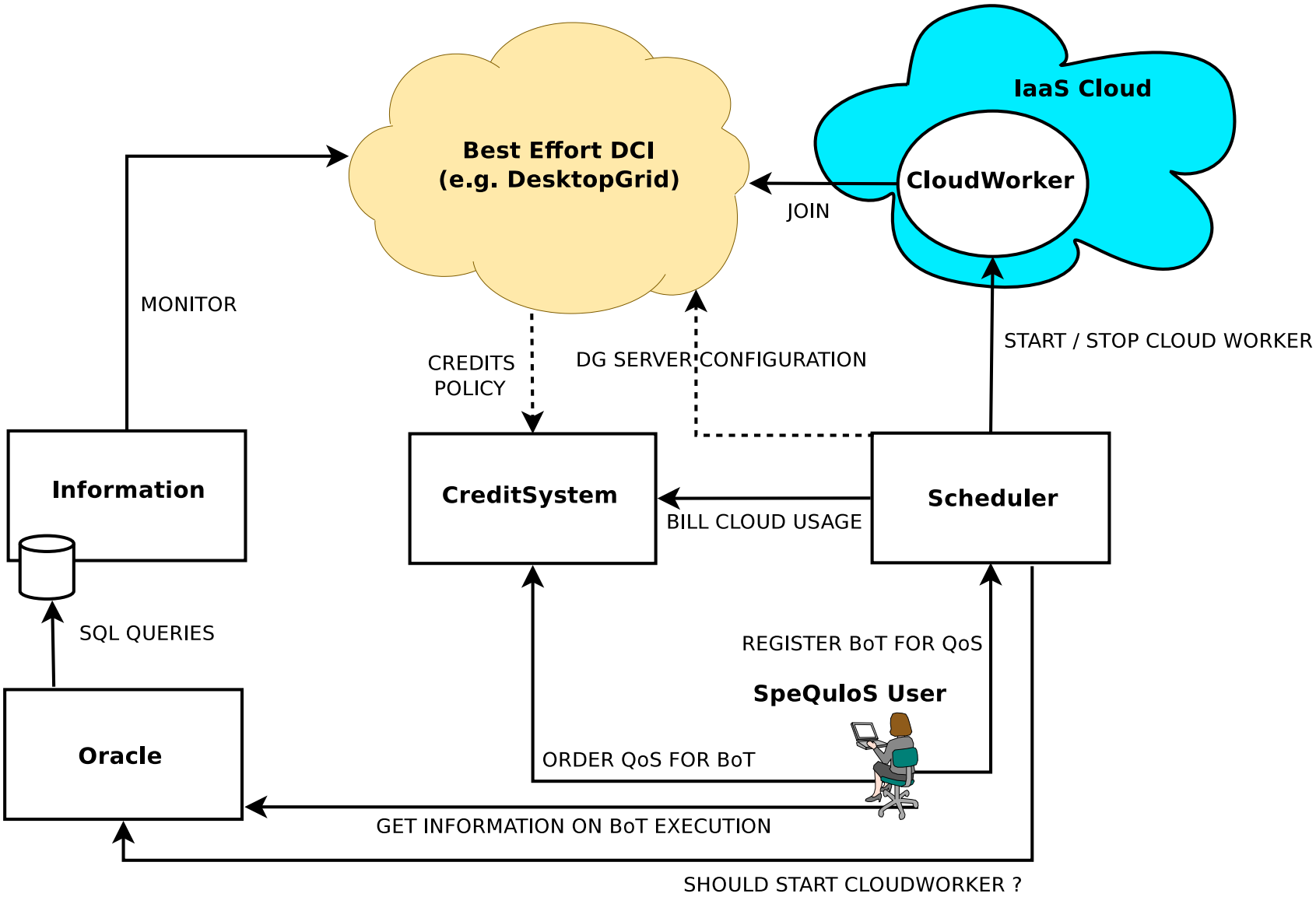
Plan

The SpeQuloS Framework

- Objectives:
  - Allow users to express QoS needs for their BoT
  - Provision resources from Cloud to satisfy these needs
- Needs:
  - Monitor iInfrastructure activity & BoT execution
  - When & How many Cloud resources to provision
  - Instanciate Cloud resources as Cloud Workers and manage it
  - Account and arbitrate Cloud usage



Overview of the SpeQuloS framework



Implementation details

DG Middle-ware

- Middle-ware supported: BOINC, XtremWeb-HEP (XWHEP)
- Modifications to make sure that workers deployed in the Cloud only process specified BoT.

$\rightarrow$ XWHEP version $\geq$ 7.3.0 & patch for BOINC server

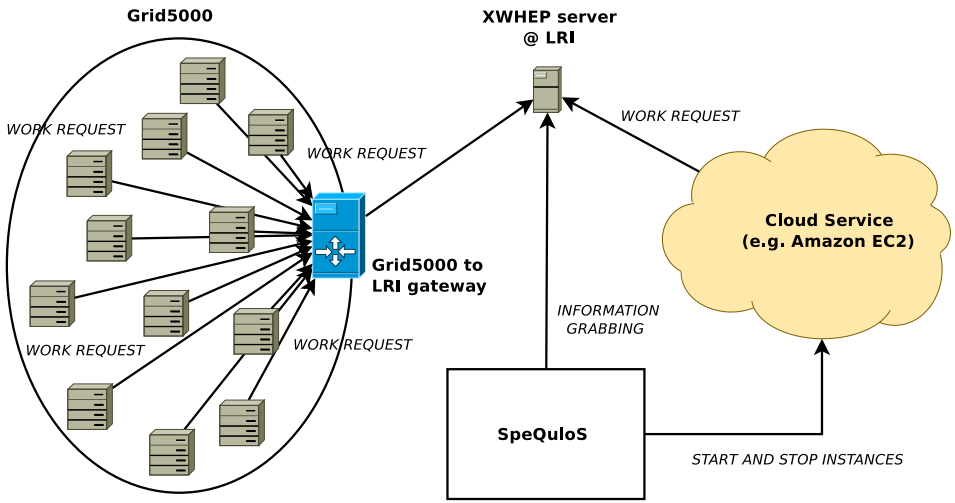
Starting Workers on the Cloud

- For each Clouds, VM images are built with DG workers software
- Cloud Workers started using libcloud and ssh to configure them
- Clouds supported : OpenNebula, EC2, Eucalyptus and we added G5K

Plan

Experimental Testbed Using Grid'5000

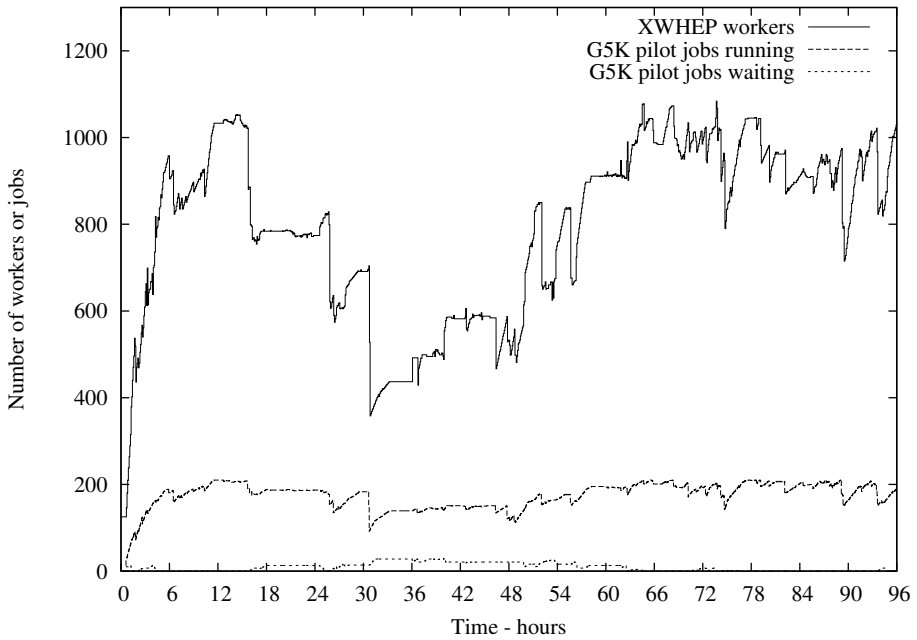
- XWHEP Desktop Grid server @ LRI, connected to EGEE
- A gateway to allow G5K nodes to connect to the XWHEP server.
- A set of XWHEP worker nodes, executed on G5K nodes.
  - A G5K job is a pilot job running several XWHEP workers (1 per core)
  - No specific environment needed
  - G5K jobs submitted in best effort queue $\rightarrow$ Variable amount of resources, unpredictable and unannounced node departure.
- SpeQuloS monitors the XWHEP server and start Cloud resources from Amazon EC2.



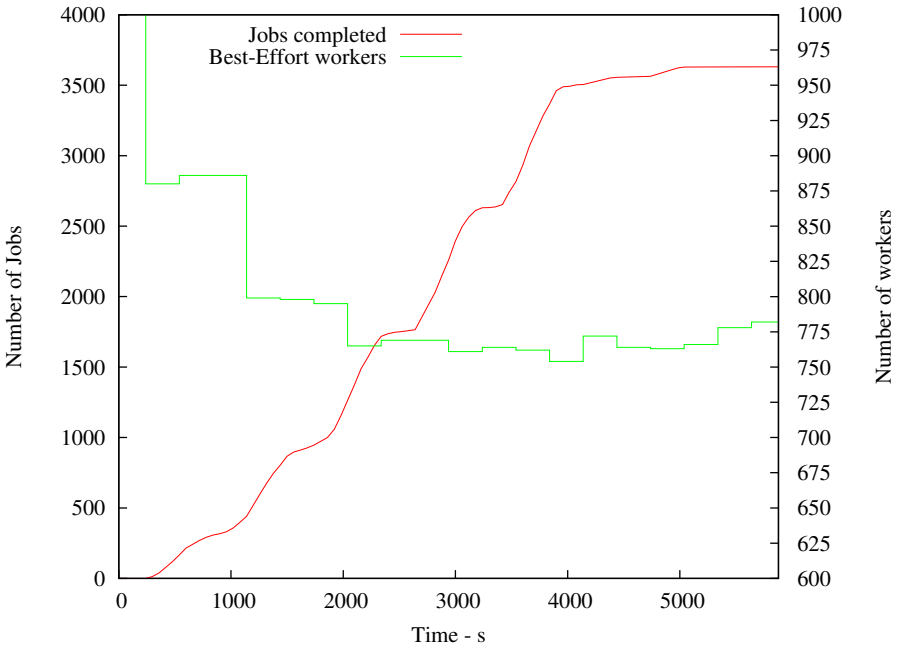
Grid'5000 resources utilization

Deployment on seven G5K sites, running:

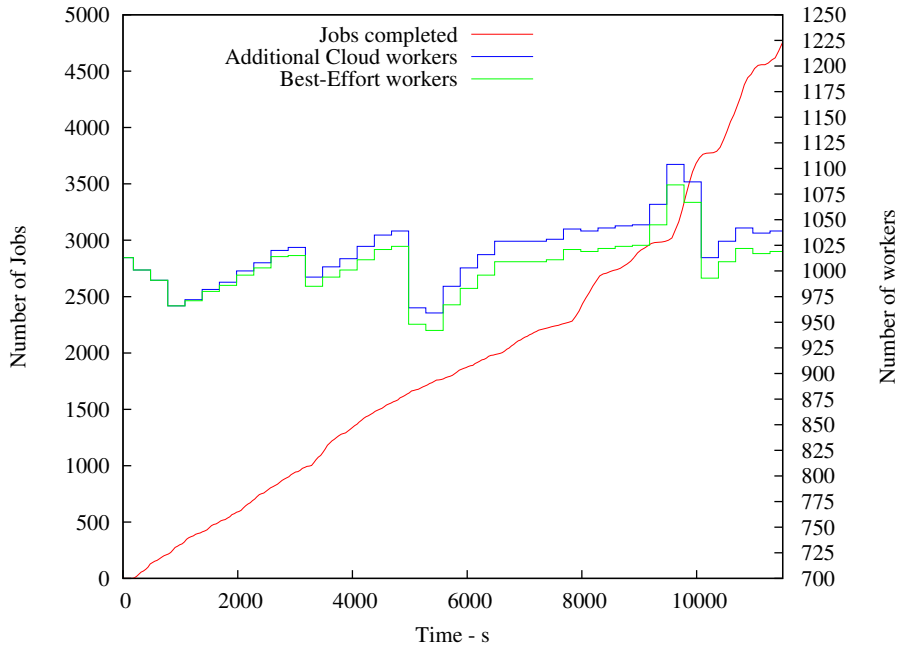
max_#pjobs ← 30
max_#pjobs_waiting ← 7
while true **do**
  if *current_#pjobs* < *max_#pjobs* **then**
    if *current_#pjobs_waiting* < *max_#pjobs_waiting* **then**
      "Submit start_XWHEP_workers to one Grid5000 node in besteffort queue"
    else
      "Too many pilot jobs waiting"
    end if
  else
    "Maximum number of pilot jobs reached"
  end if
  sleep(15 minutes)
end while



First results



SpeQuloS not used



SpeQuloS used

→ the tail has disappeared ! But not enough experiments to conclude.

Plan

Using Grid'5000 as a Cloud

- For experimentations, it is difficult to use Clouds is :
  - Using public Clouds like EC2 is costly
  - Using private Clouds:
    - Is complex to deploy and maintain
    - Needs a lot of hardware to be useful.
- Grid'5000 has most of IaaS Cloud features
  - On-demand resources availability
  - User-driven execution environment using deployment
  - Remote access through API
- Idea: Conduct SpeQuloS experiments with Grid5000 as a Cloud.

The Grid'5000 libcloud driver

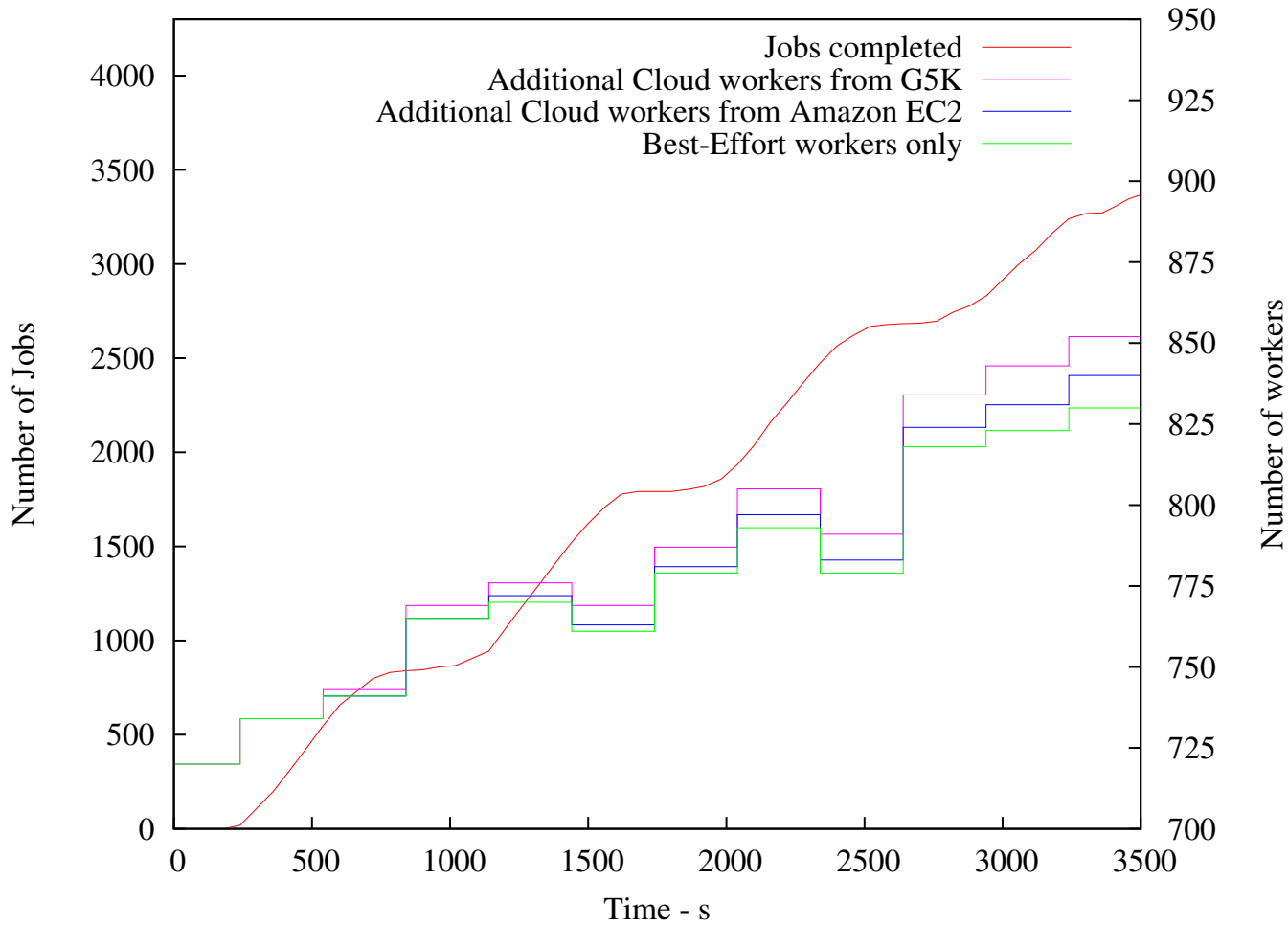
- libcloud:
  - Python API used in SpeQuloS
  - Common interface to various Cloud technologies
  - Support: Amazon EC2, Eucalytpus, OpenNebula, RackSpace, Nimbus, ...
- We propose a new "driver" for libcloud: Grid'5000 driver
  - Using the Grid'5000 API
  - Implement standard libcloud features:

libcloud feature	**Grid5000 implementation**
Start/Stop instance	Interactive job submission
List Node Sizes	List available nodes
List Disk Image	List available environment to deploy

- Still few work to be completed

Experimentation results

- Grid5000 is used as Best Effort DCI
- Both EC2 and Grid5000 are used as Cloud resources
- Grid5000 as a Cloud is hosted on the Rennes site



Plan

Conclusion

On-going works regarding SpeQuloS

- Improving SpeQuloS basics:
  - improve real-time estimation of completion time (moving average)
  - improve the detection of the tails by fitting statistical distribution to BoT execution archive
  - improve scheduling of Cloud resources, i.e. when and how many Cloud Workers to start ?

→ More G5K experiments to validate SpeQuloS

Remarks on Grid5000 utilization

- Grids besteffort queue as a Best Effort DCI
- Grid5000 can be used as a Cloud
  - Without additional virtualization
  - Not as flexible as real Cloud (deployment, node size, isolated network)
  - libcloud driver release will be announced on the mailing list