Simulation of Price-sensitive Resource Brokering and the Hybrid Pricing Model with DGAS-Sim

Rosario M. Piro

Andrea Guarise, Albert Werbrouck

Istituto Nazionale di Fisica Nucleare (INFN), Turin/Italy

WETICE (ETNGRID) 2004, Modena, Italy

piro@to.infn.it

Grid Computing – some difficulties

- The grid is dynamic in nature: Resources are added/removed or are not always available (e.g. due to network problems or system failures).
- Resources usually belong to different administrative domains with different policies for access and utilization.
- The utilization of the single resources (and hence their workload) is unpredictable:
  - scheduling algorithms usually used for parallel computing (e.g. for cluster computing) are less adequate;
  - grid computing requires more flexible "just-in-time" scheduling for the submitted jobs

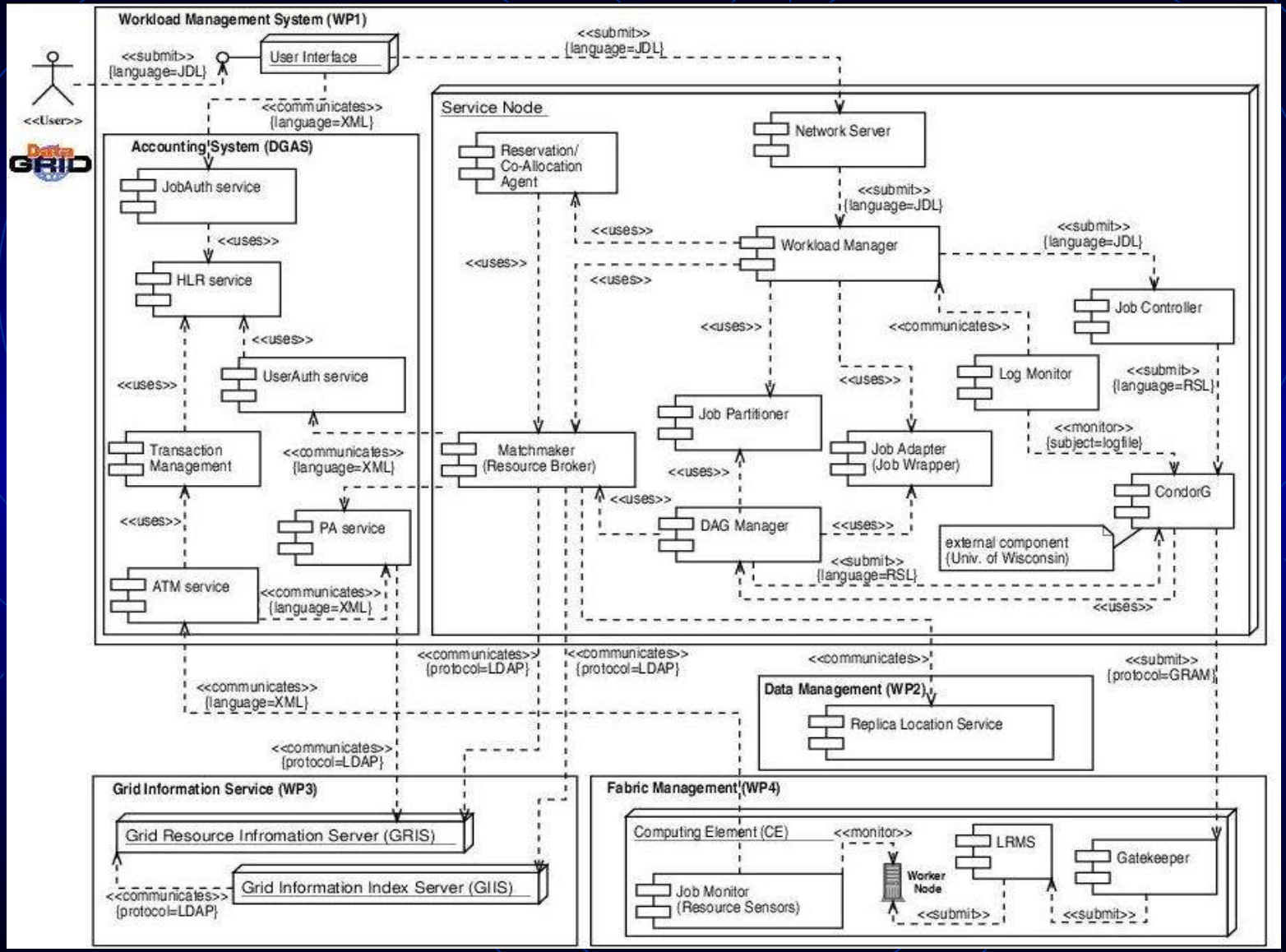
piro@to.infn.it

Preamble

- research work done within the European DataGrid Project (EDG), that is now being continued as EGEE project (Enabling Grids for E-science in Europe).
- the simulations are based on the architecture of the EDG Workload Management System, adapted to support economic brokering. In the EGEE architecture, resource pricing issues will probably be included in a more complex policy framework.

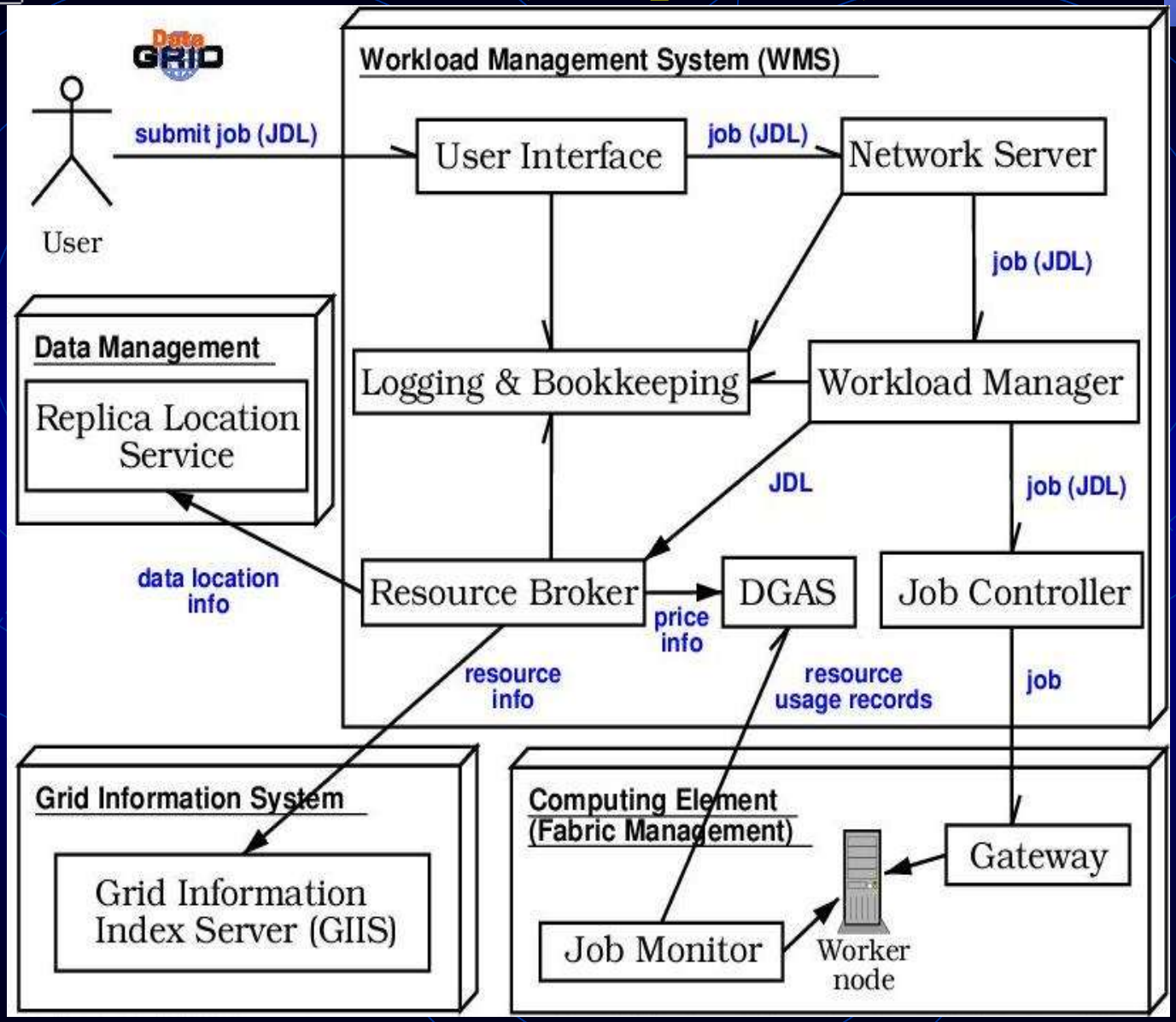
piro@to.infn.it

EDG Workload Management System



piro@to.infn.it

WMS (simplified)



piro@to.infn.it

Motivation(1): why grid accounting?

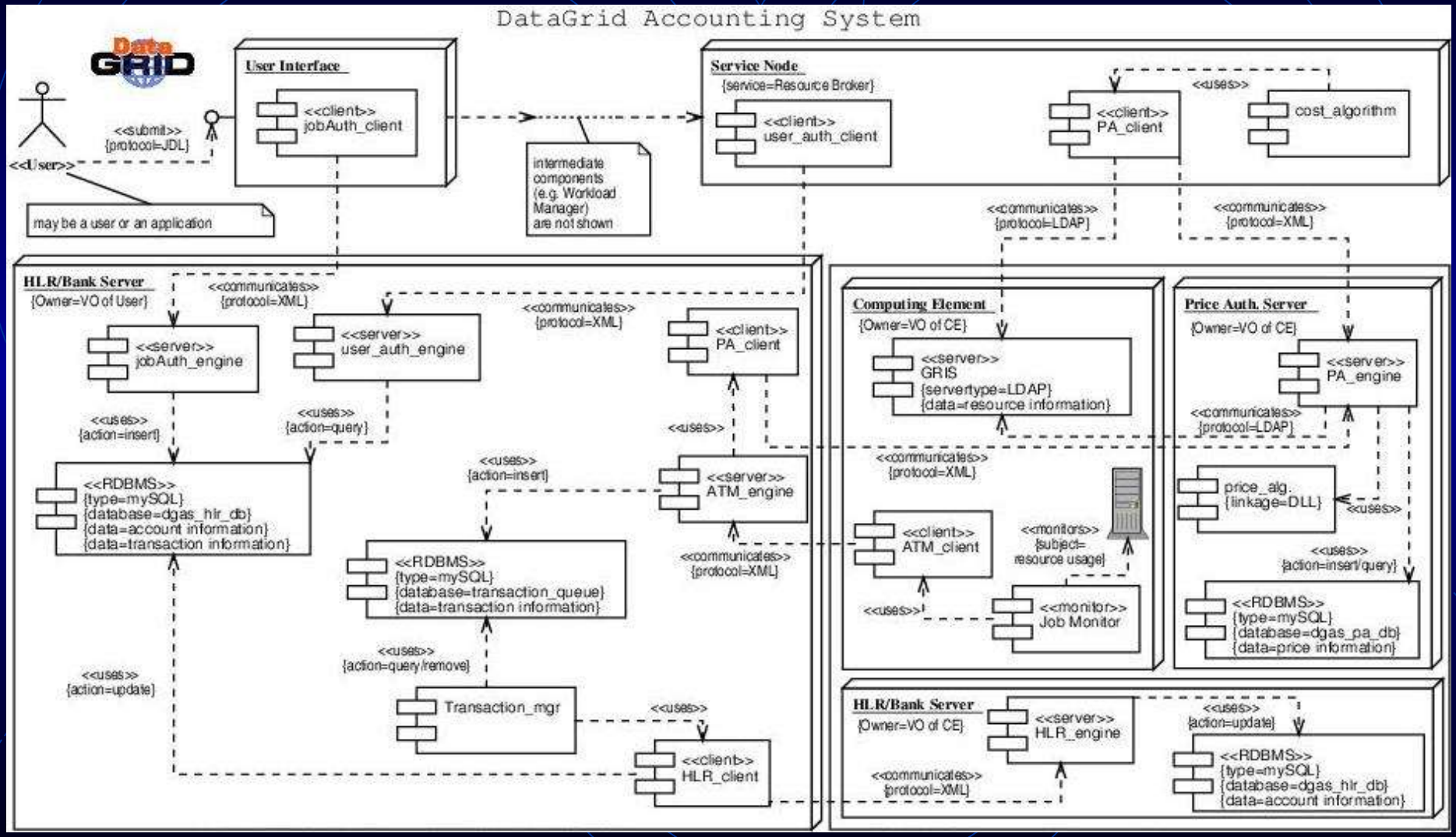
- keep track of the resource usage by the single grid users
- avoid abuses of the resources (e.g. due to limited budgets)

- brokering (resource selection) based on economic principles may help the Workload Management System to balance the workload.

→ facilitate the fair and balanced exchange of computing resources

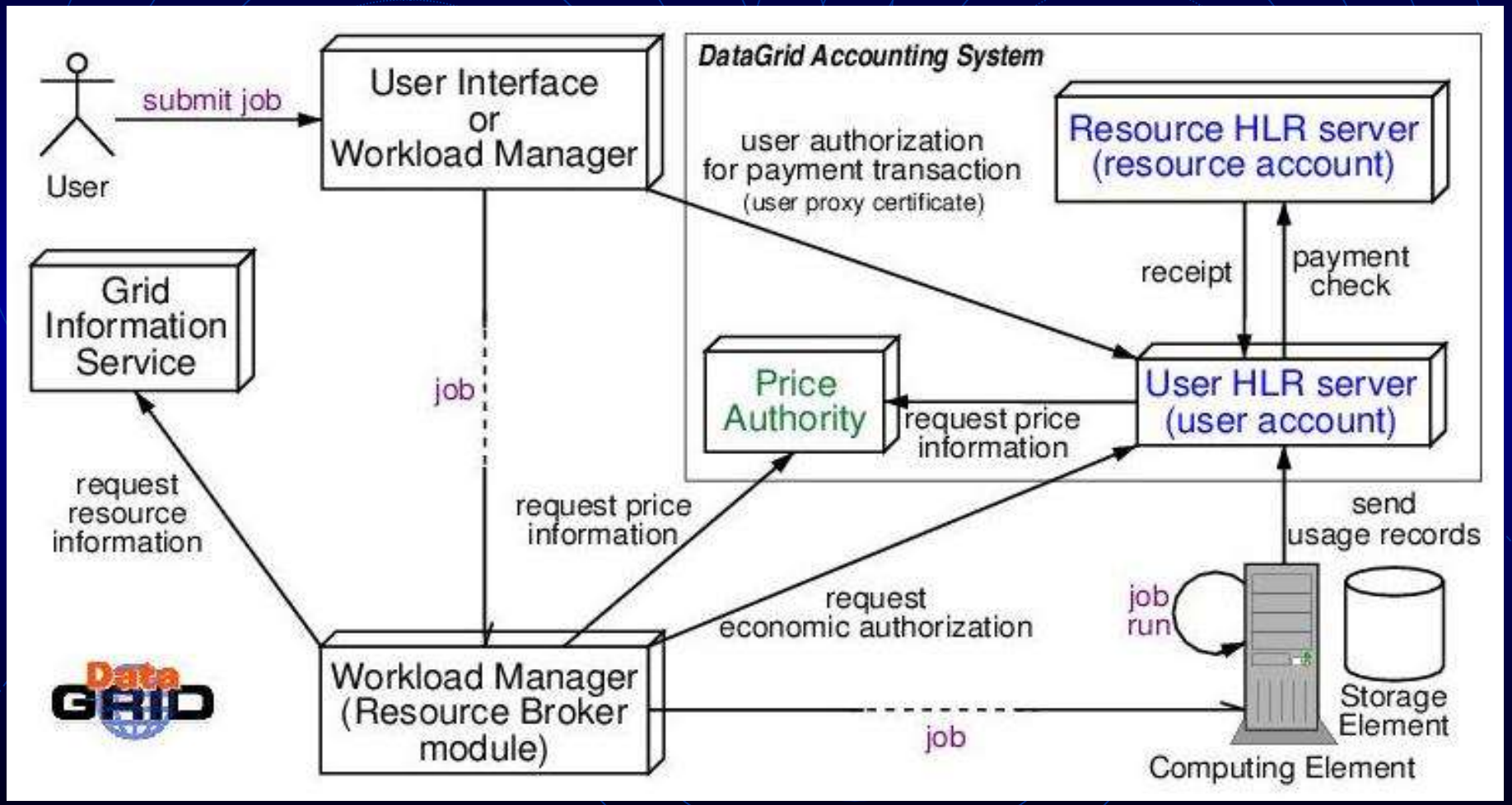
piro@to.infn.it

DataGrid Accounting System (DGAS)



piro@to.infn.it

Interaction between DGAS and other EDG components (simplified)



piro@to.infn.it

Motivation(2): why economic brokering?

- all grid entities (users, computing resources, storage resources,..) spend or earn "virtual credits" (GridCredits) for their grid related activities.
- resource prices are determined *dynamically*. An *economic feedback* may help the Resource Broker to select less loaded (possibly idle) resources for the submitted jobs, if the prices reflect the resources' workloads.
- allows for a *decentralized* optimization of the distribution of jobs (offering the *scalibility* desired for grid computing).
- may bring about *market equilibrium* in which the requested computing power equals the computing power furnished by the resources.

piro@to.infn.it

Hybrid Pricing Model (HPM)

- pricing algorithm (price computed for each single resource):

$$price = P_0 + \Delta P \frac{W - \frac{1}{2}W_{max}}{\frac{1}{2}W_{max}} \qquad 0 \leq W \leq W_{max}$$

- *static parameters* : base price P_0, variation limit ΔP and maximum Queue Wait Time W_{max}

- *dynamic parameter*: current Queue Wait Time W (estimated)

➡ price dynamically varies between $P_0 - \Delta P$ (empty queue) and $P_0 + \Delta P$ (full queue) *as a function of the resource's Queue Wait Time*

piro@to.infn.it

Computational Energy

- Prices are expressed in Grid Credits per Unit of Computational Energy (GC/UCE)
- We define the *computational energy* "consumed" by a job as the product of a *performace factor* or *power p* (e.g. benchmark for processing power) and the *amount of usage u* (e.g. processing time).
- The consumption of computational energy by a job is ideally independent of the resource that is executing it.
- For the presented simulation:
  - assume the clock speed to be a sufficiently good measure for processing power
  - 1 UCE = 1 MHz × min.

piro@to.infn.it

Advantages and disadvantages of the HPM

- simplicity, low computational overhead and low communication overhead (only minimal information required).
- variation limits prevent the degeneration of resource prices (e.g. due to inflation).
- requires definition of maximum QWTs for predefined queue classes (e.g. short, medium, long ..)
- requires a reasonable estimate of current QWTs (and thus an estimation of job requirements in terms of computational energy).

  => possible: estimation by user or based on historical data for similar jobs (W.Smith, V.Taylor and I.Foster, 1999; H.Li, D.Groep, J.Templon and L.Wolters, 2004).

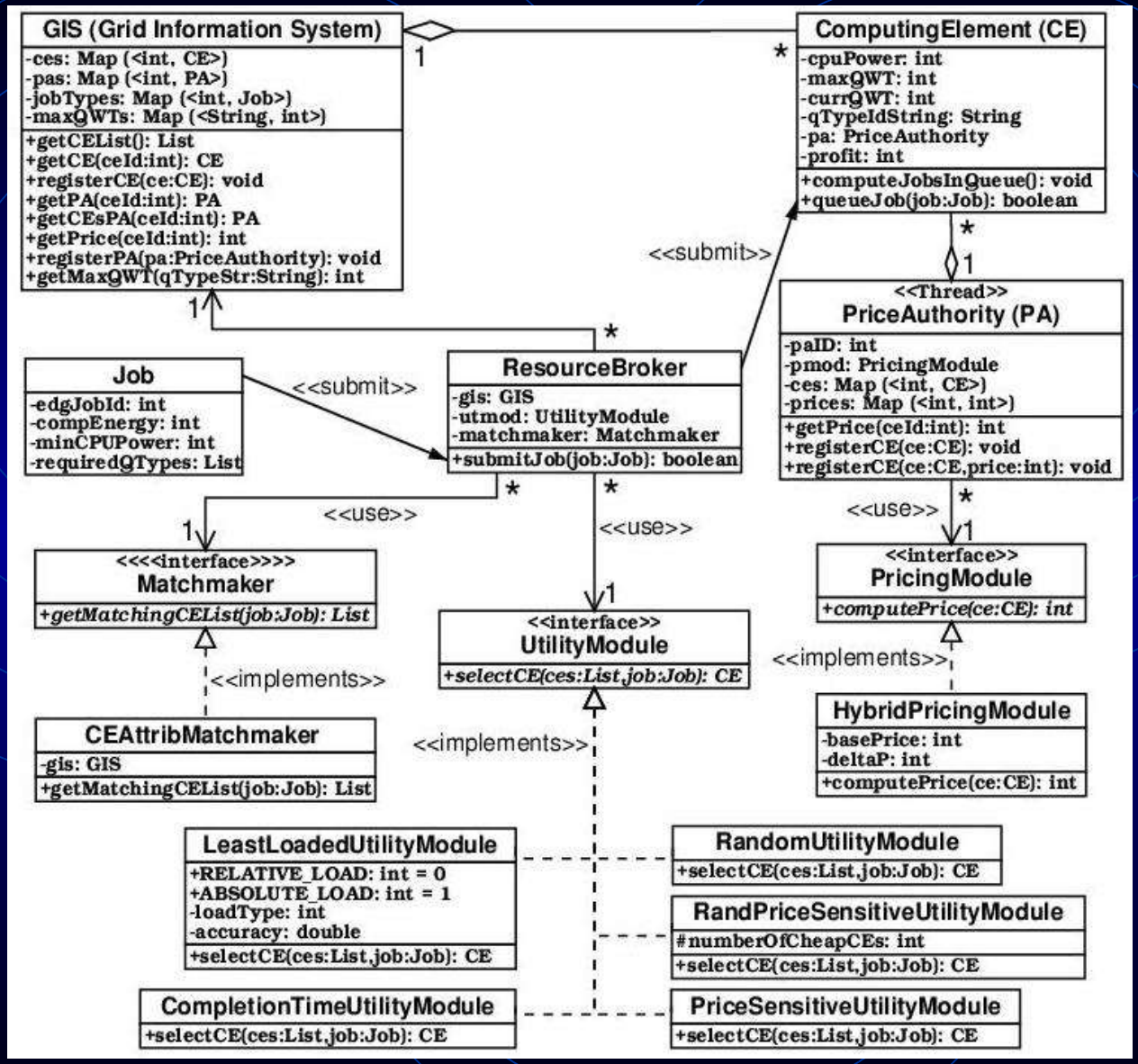
piro@to.infn.it

DGAS-Sim(ulator)

- modular and expansible tool to *simulate the components involved in the distribution of the workload* (Resource Broker, Grid Information System, Price Authorities, Computing Elements).
- interfaces for brokering strategies and pricing algorithms allow the simulation of different behaviors of Resource Brokers and Price Authorities to *study their impact on workload balancing*.
- *a configuration file* allows to specify an arbitrary number of Computing Elements (resources) and Price Authorities, the interval between price adjustments, different job types and their consumption of computational energy and their requirements in terms of computing power, etc.
- simulation of single-processor resources with one FCFS queue (we focus on load balancing *among* grid resources).
- data location and network traffic are not considered.

piro@to.infn.it

DGAS-Sim: architecture



piro@to.infn.it

Brokering strategies

- ***price-sensitive***: prices are the only criterion for resource selection; the Resource Broker chooses the resource (Computing Element) with the *lowest price*; other factors such as data access times and network bandwidths are not considered.

For comparison:

- ***random*** (no optimization): jobs are submitted to *randomly chosen* resources that satisfy their requirements.
- ***completion time*** (best case): the criterion for resource selection is the given job's *completion time* (QWT + execution time); the Resource Broker chooses the resource with the *lowest completion time* (that is assumed to be exactly known).

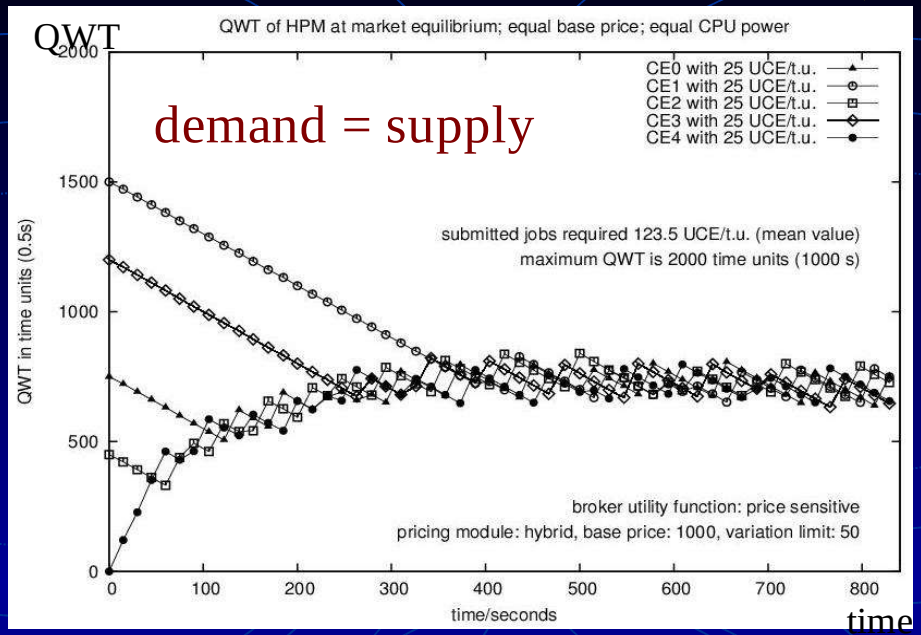
piro@to.infn.it

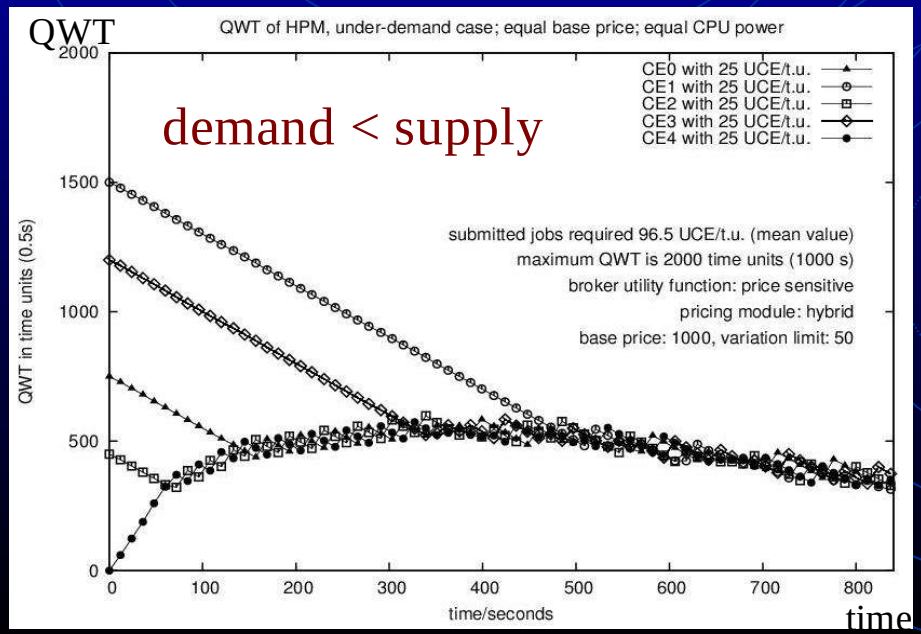
First simulation results (I)

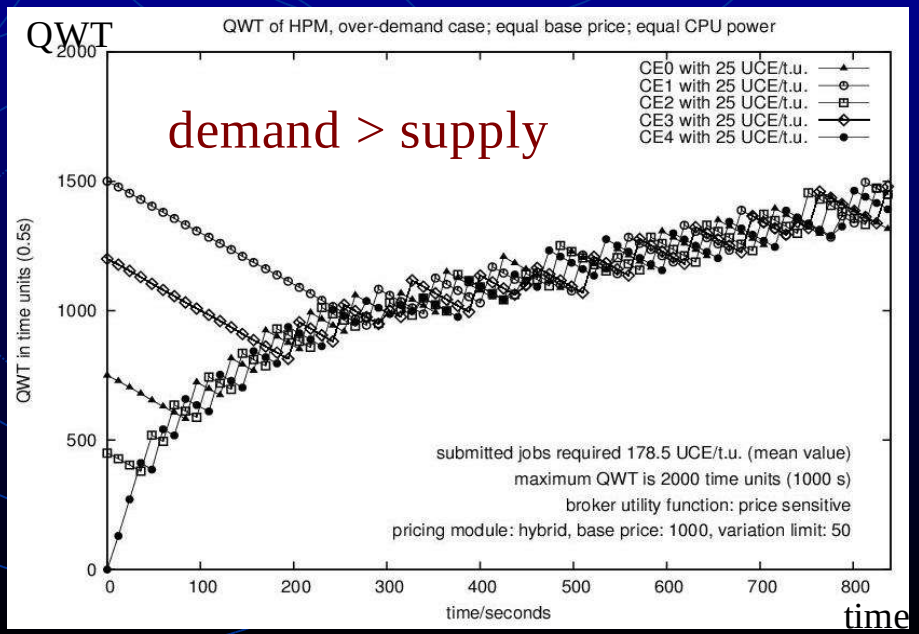
- pricing algorithm: HPM with small price variations (about ±5%); price adjustment interval of about 12 seconds; small number of resources
- brokering strategy: price-sensitive
- the queues of the resources are *initially unbalanced*
- three cases:
  - market equilibrium (demand = supply)
  - overdemand (demand > supply)
  - underdemand (demand < supply)

piro@to.infn.it

First simulation results (II)







Simulation of a small grid

- 50 Computing Elements (CEs)
  - with 5 different queue types:
    - *very short queues*: maximum QWT of 2 hours
    - *short queues*: maximum QWT of 4 hours
    - *medium queues*: maximum QWT of 10 hours
    - *long queues*: maximum QWT of 18 hours
    - *very long queues*: maximum QWT of 24 hours
  - with 5 different processing powers:
    - 2 CEs of each queue type with 800 UCE/min (~ 800 MHz)
    - 2 CEs of each queue type with 1200 UCE/min (~ 1.2 GHz)
    - 2 CEs of each queue type with 1600 UCE/min (~ 1.6 GHz)
    - 2 CEs of each queue type with 2000 UCE/min (~ 2.0 GHz)
    - 2 CEs of each queue type with 2400 UCE/min (~ 2.4 GHz)

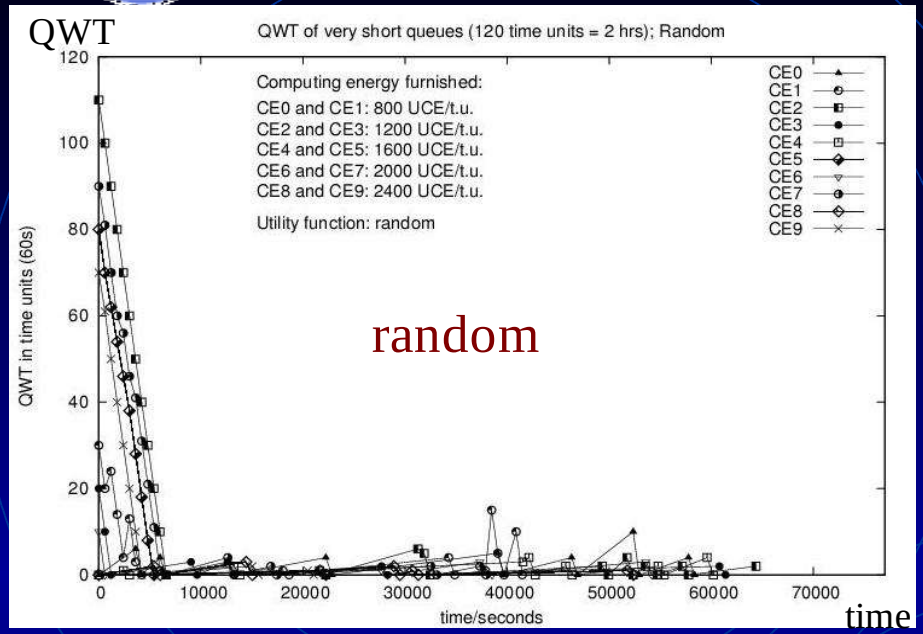
piro@to.infn.it

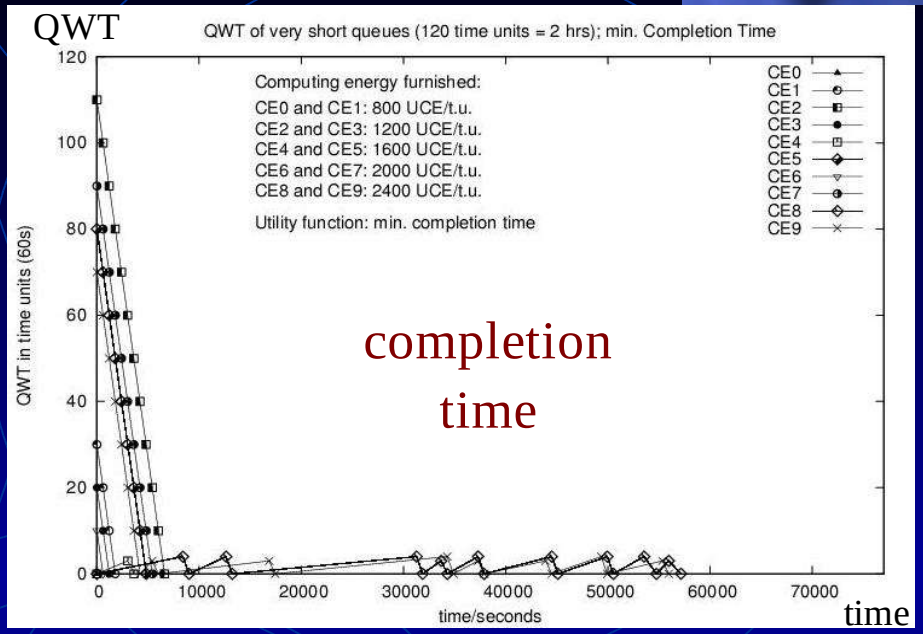
Simulation of a small grid

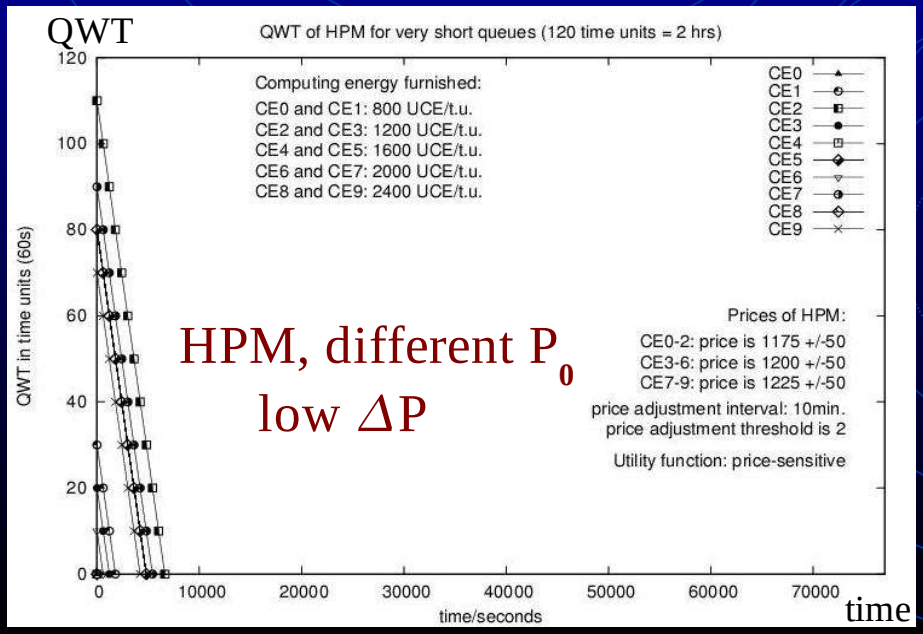
- one job submission about every 30 seconds; 20 different job types with different requirements in terms of processing power and queue types (to simulate matchmaking):
  - *underdemand for shorter queues,*
  - *overdemand for longer queues.*
- brokering strategies: price-sensitive, random, completion time
- pricing algorithm (for price-sensitive brokering): HPM with a price adjustment interval of about 10 minutes and *initially unbalanced* queues
  - *Sim. run 1*: base prices P_0 between 775 and 1225 GC/UCE; low price variation limit ΔP of 50 GC/UCE (=> price domains do not overlap significantly)
  - *Sim. run 2*: common base price of P_0 = 1000 GC/UCE for all CEs; price variation limit ΔP of 50 GC/UCE

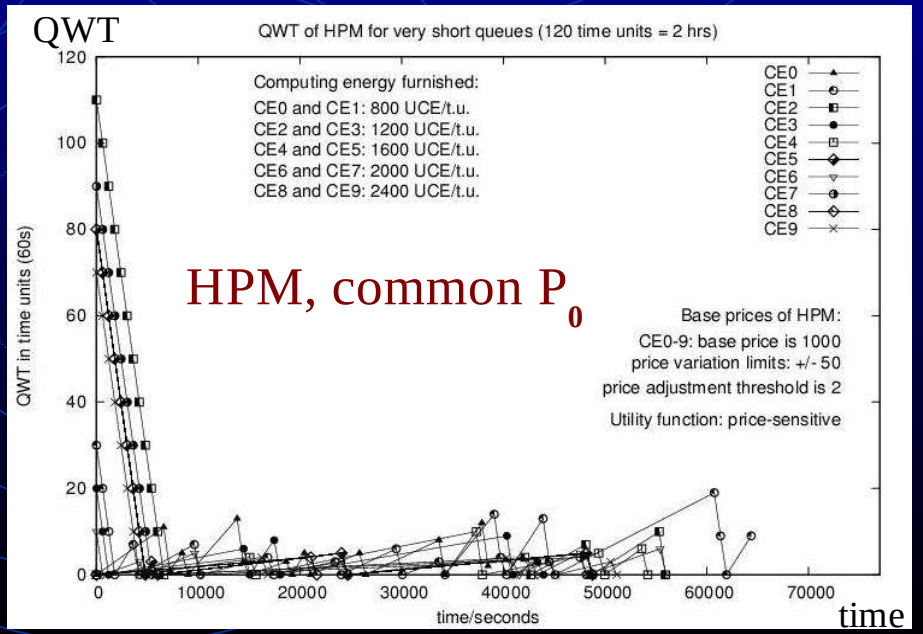
piro@to.infn.it

Results (1): QWT, very short queues

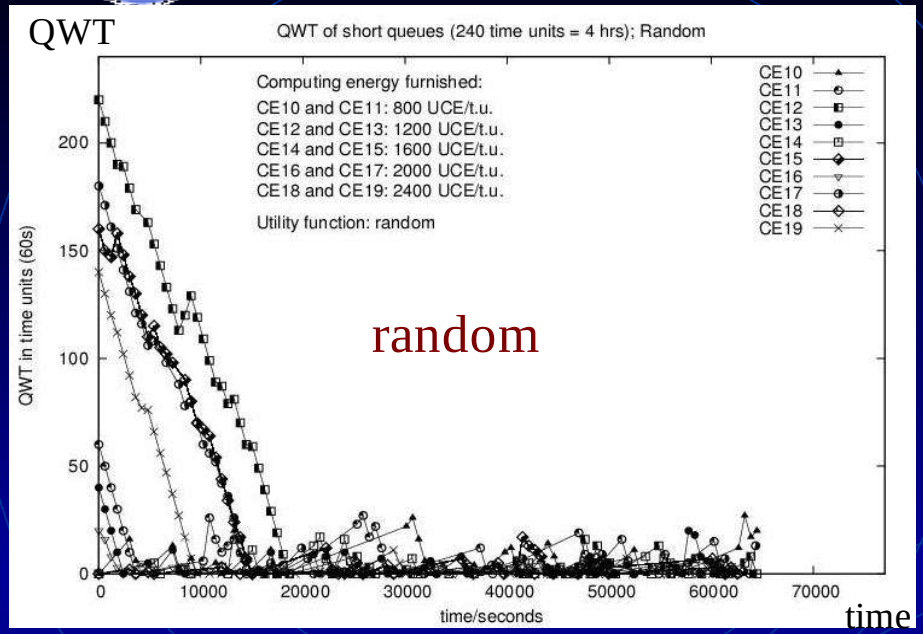


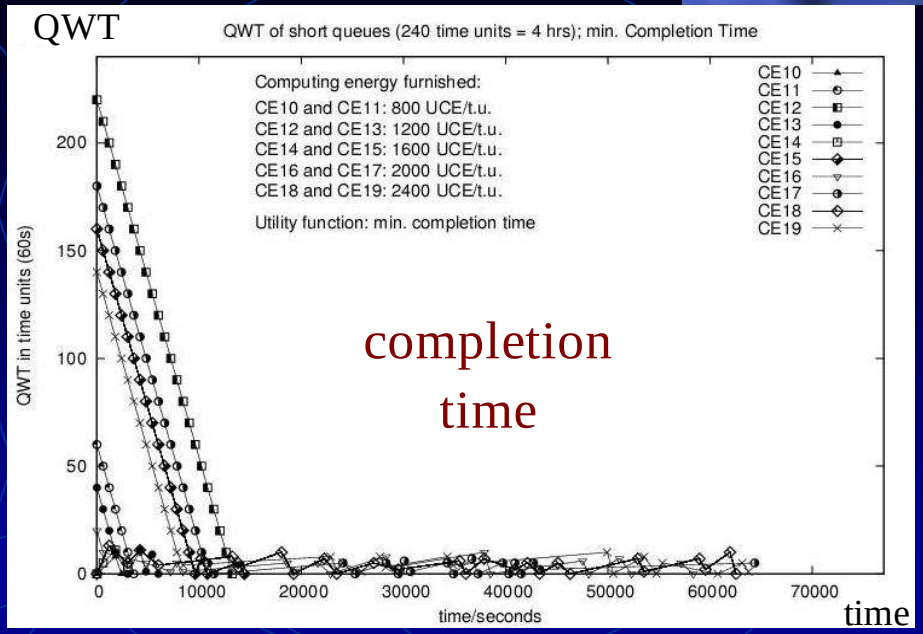


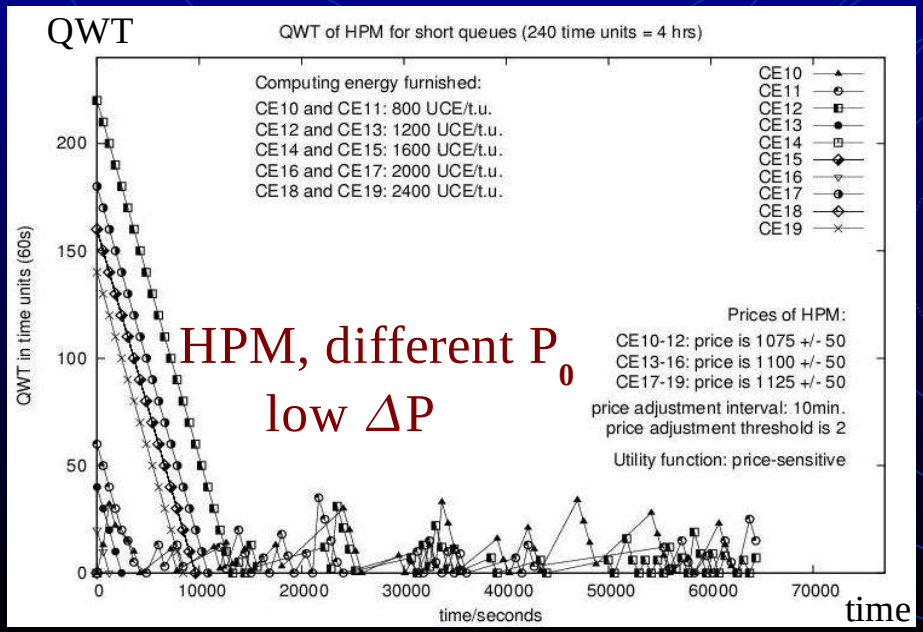


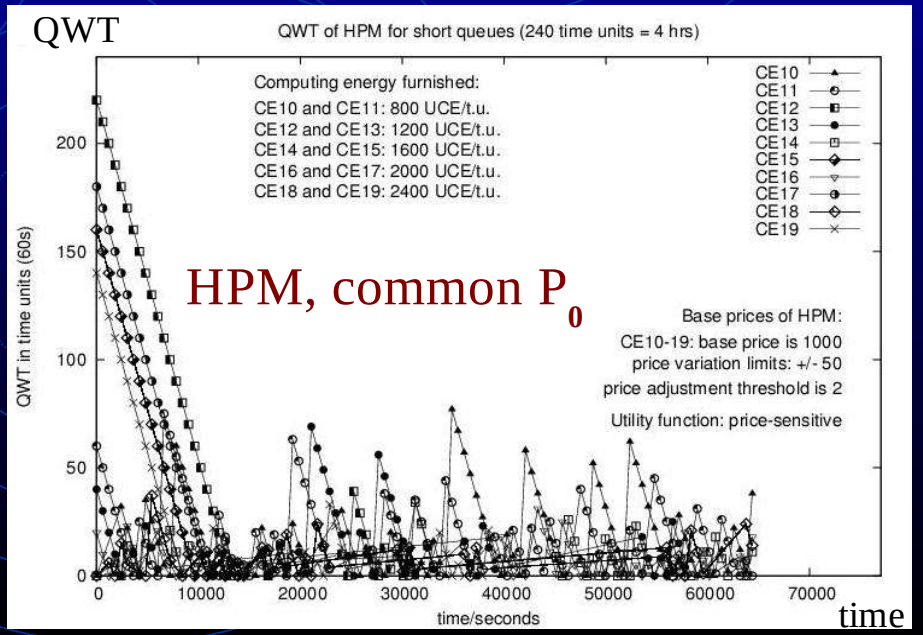


Results (2): QWT, short queues

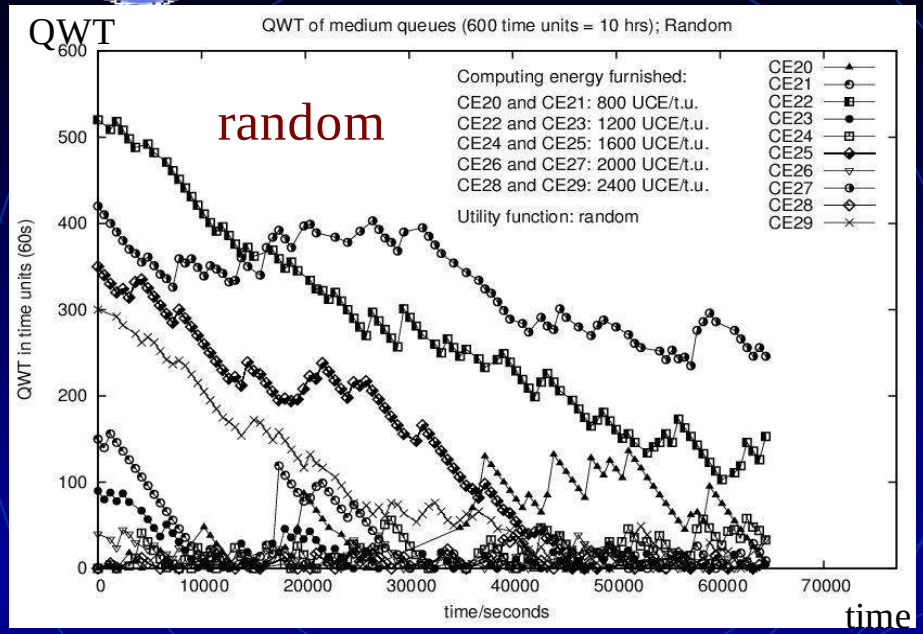


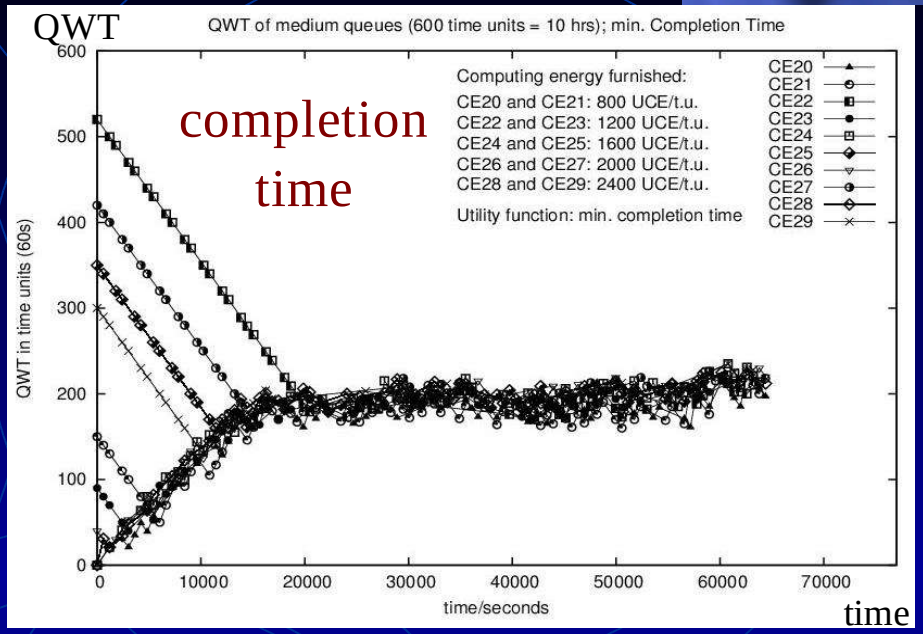


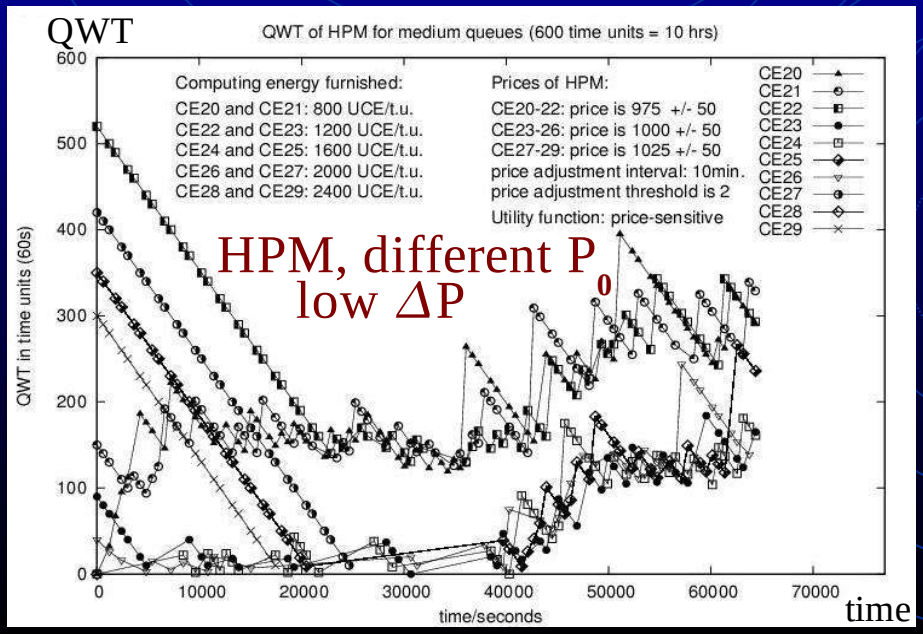


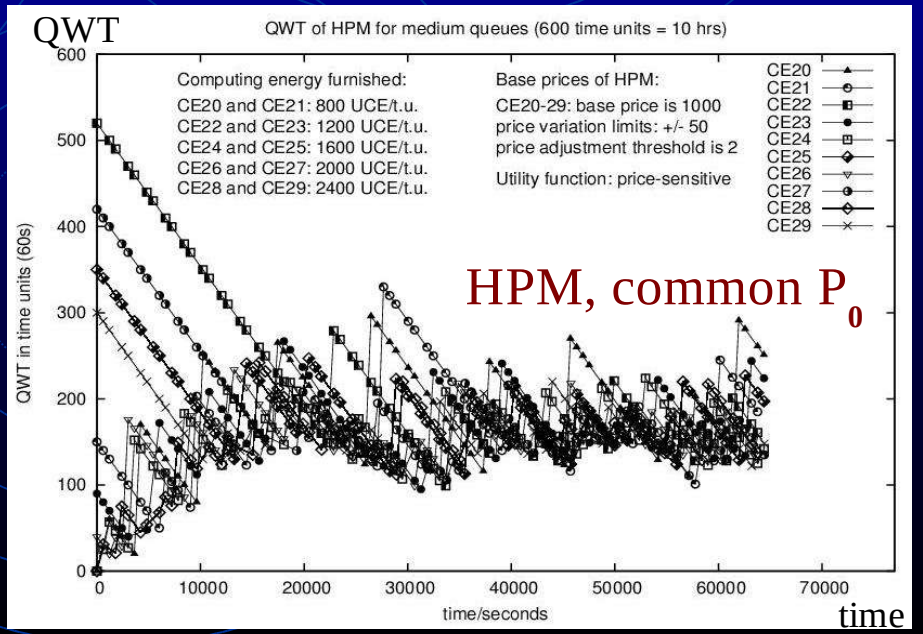


Results (3): QWT, medium queues

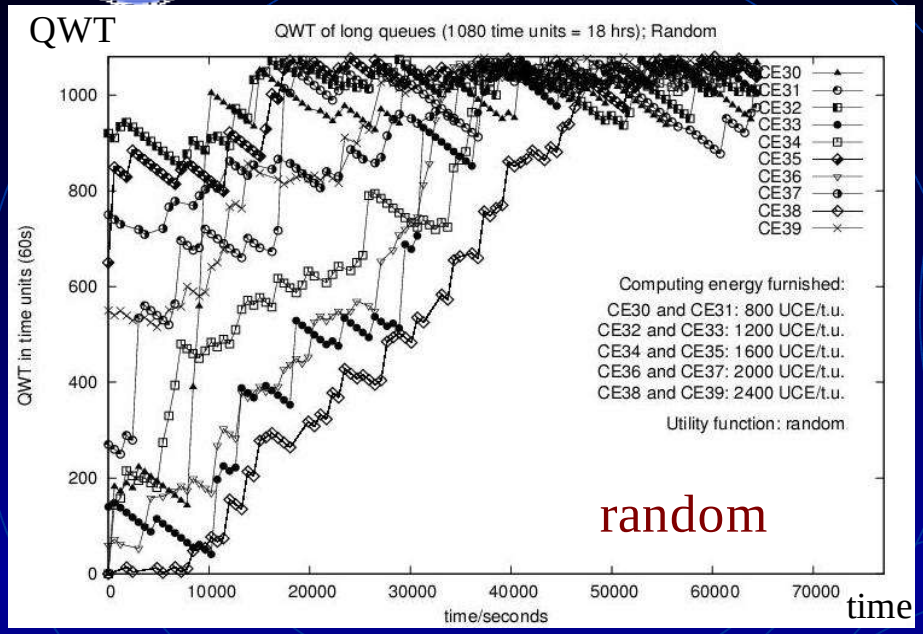


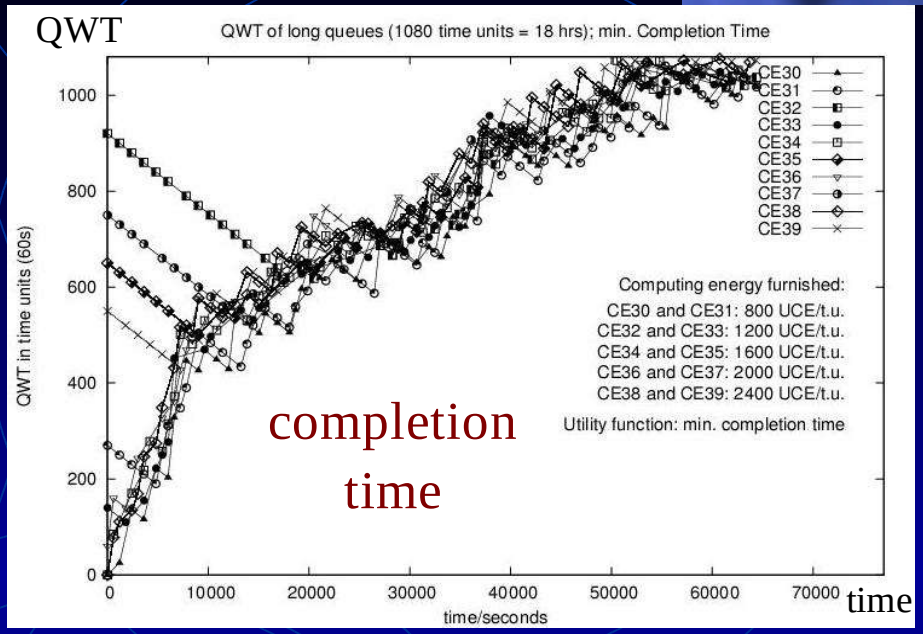


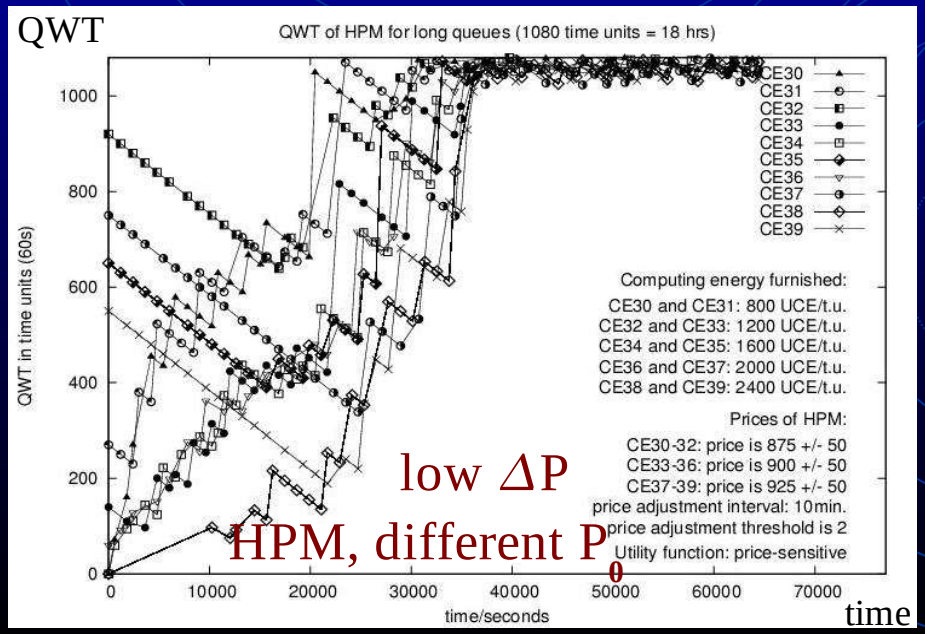


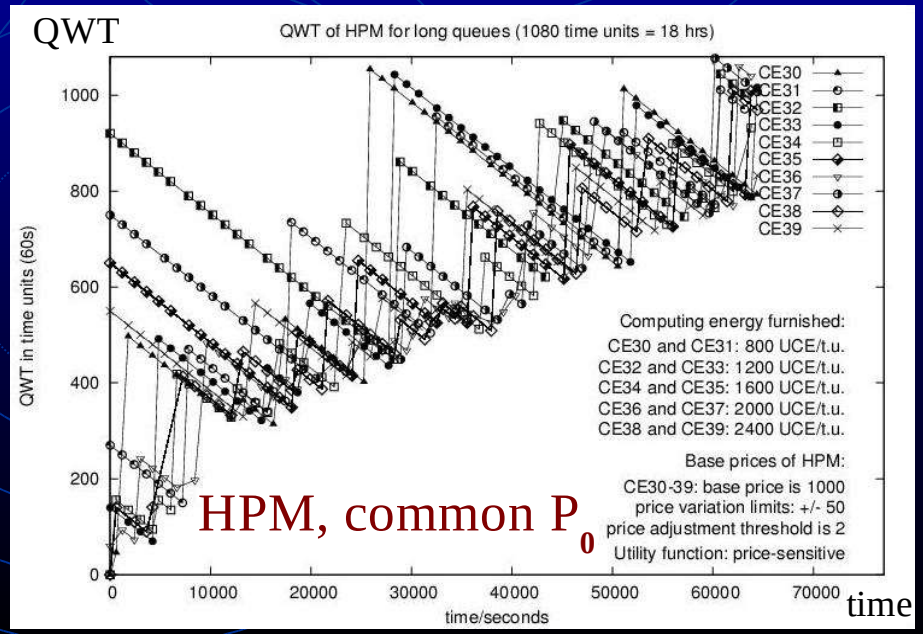


Results (4): QWT, long queues

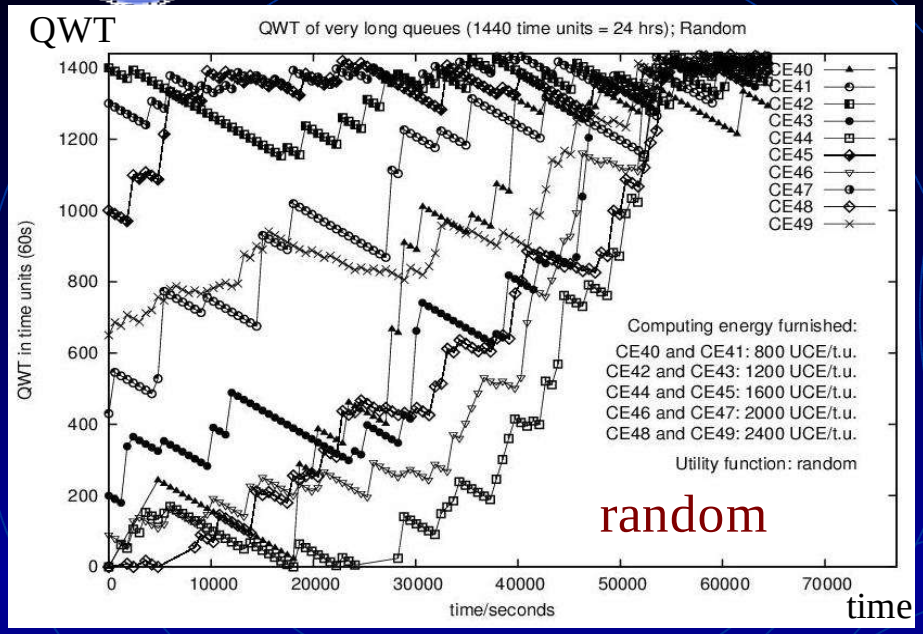


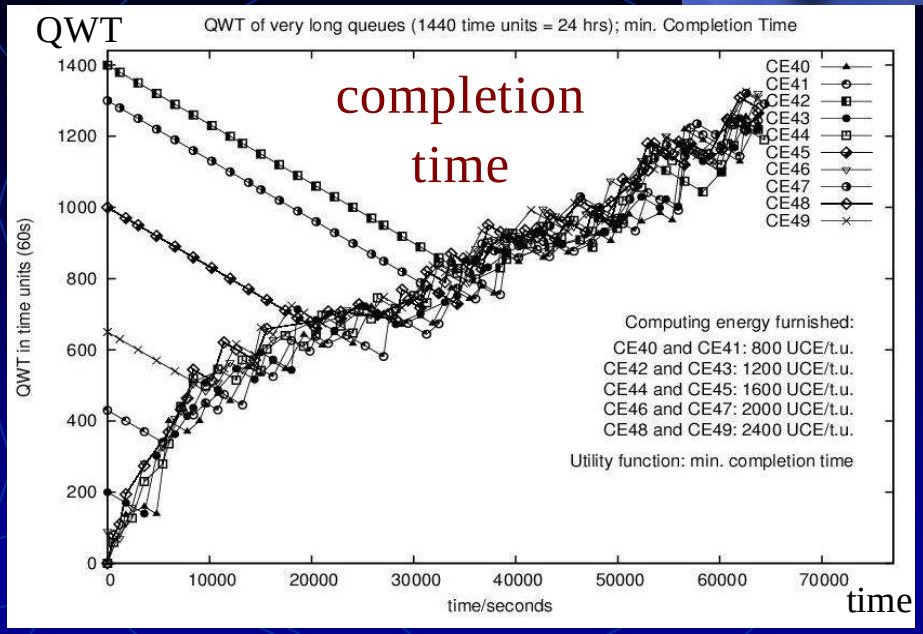


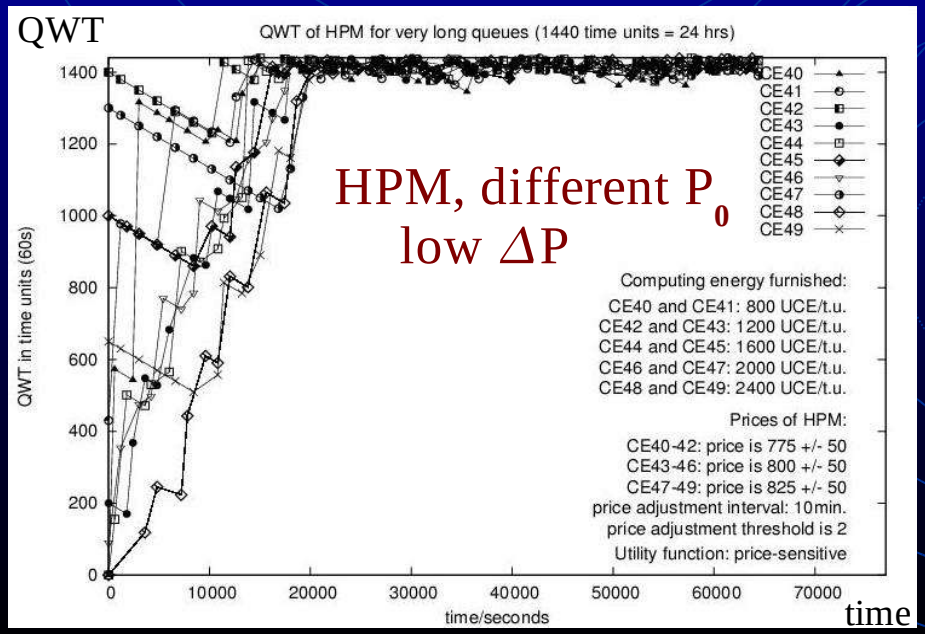


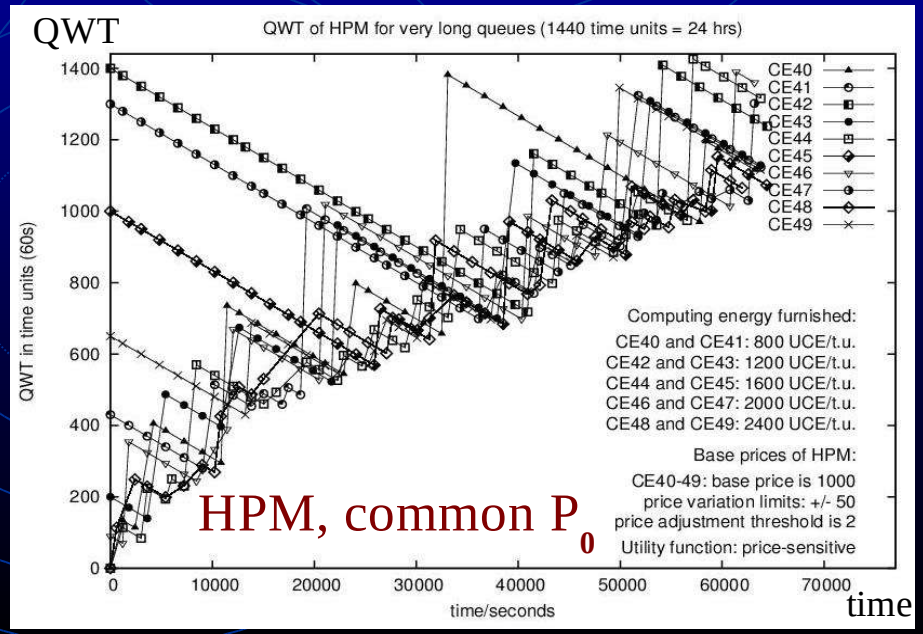


Results (5): QWT, very long queues

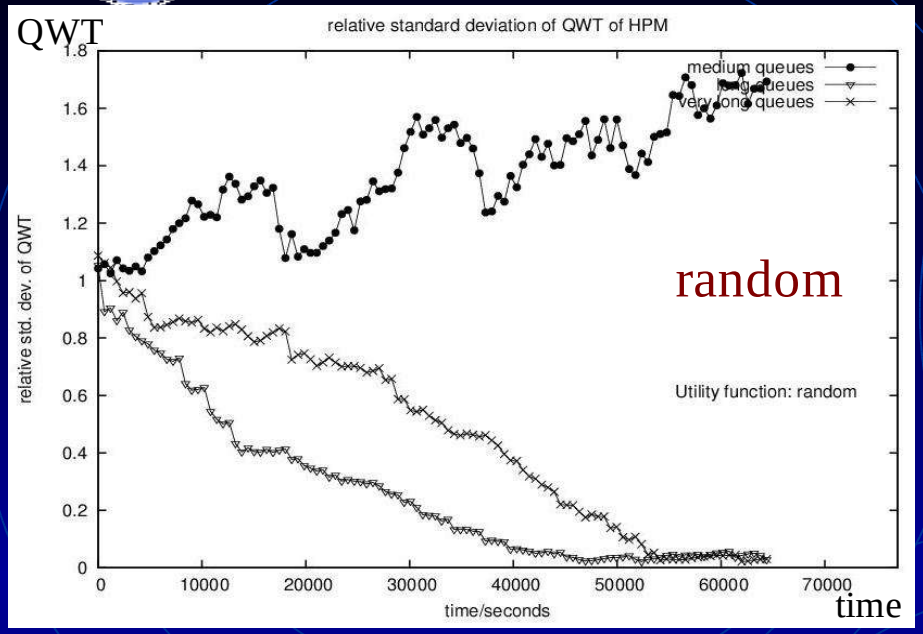


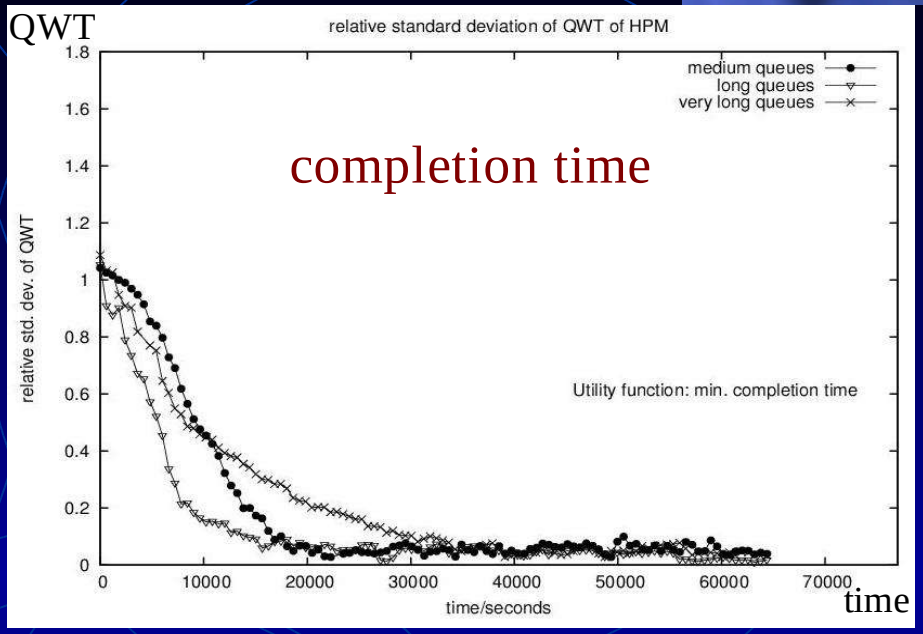


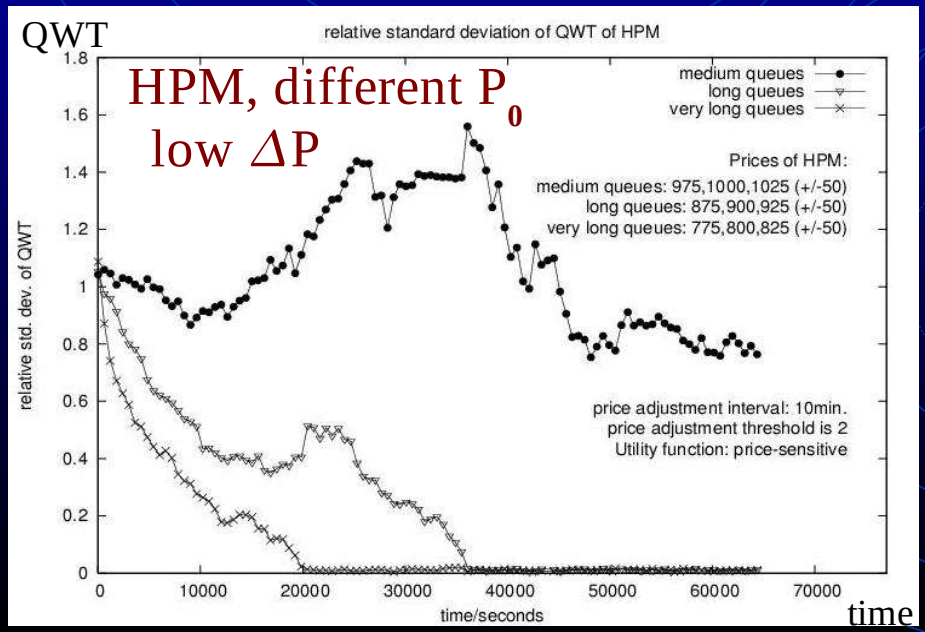


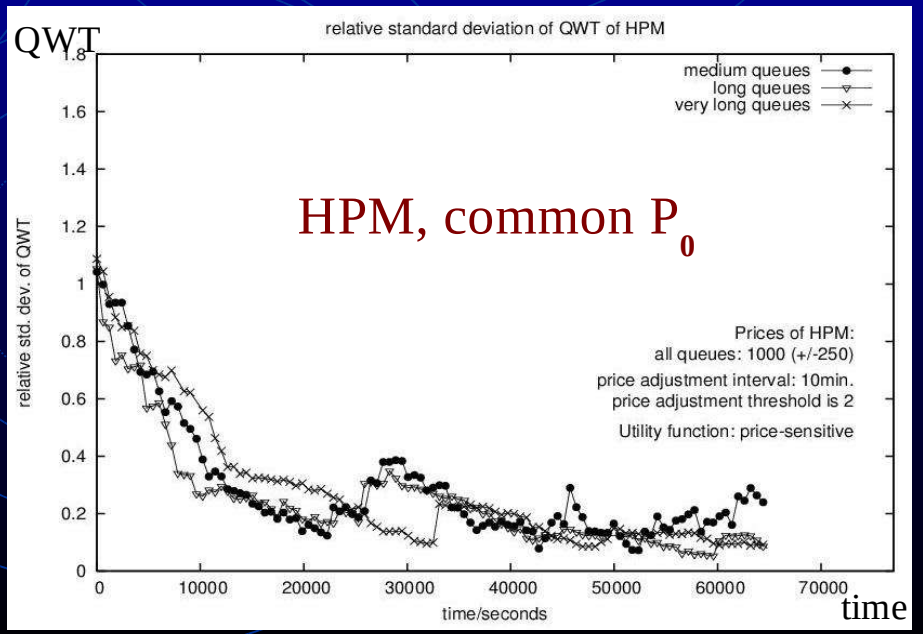


Results (6): relative standard deviation







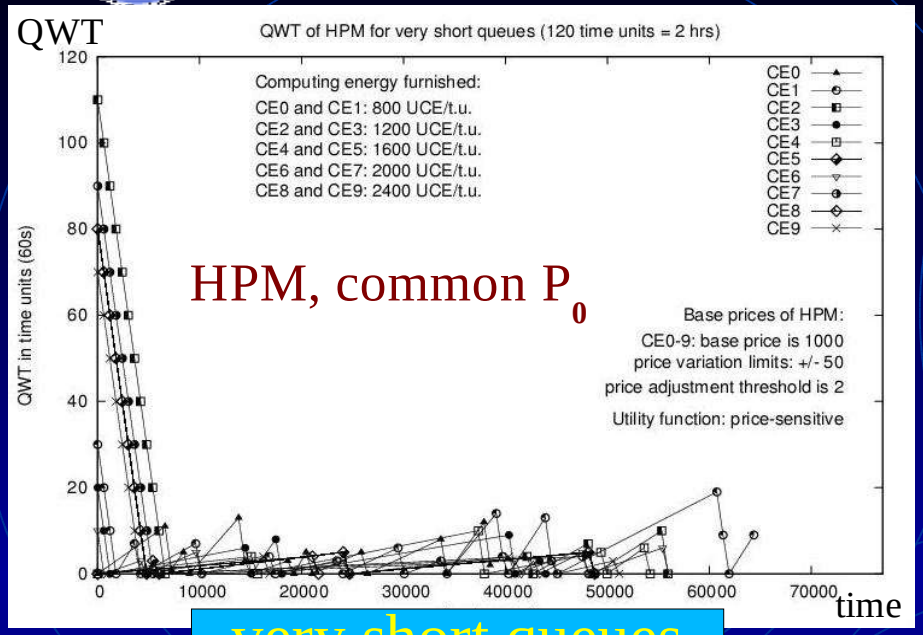


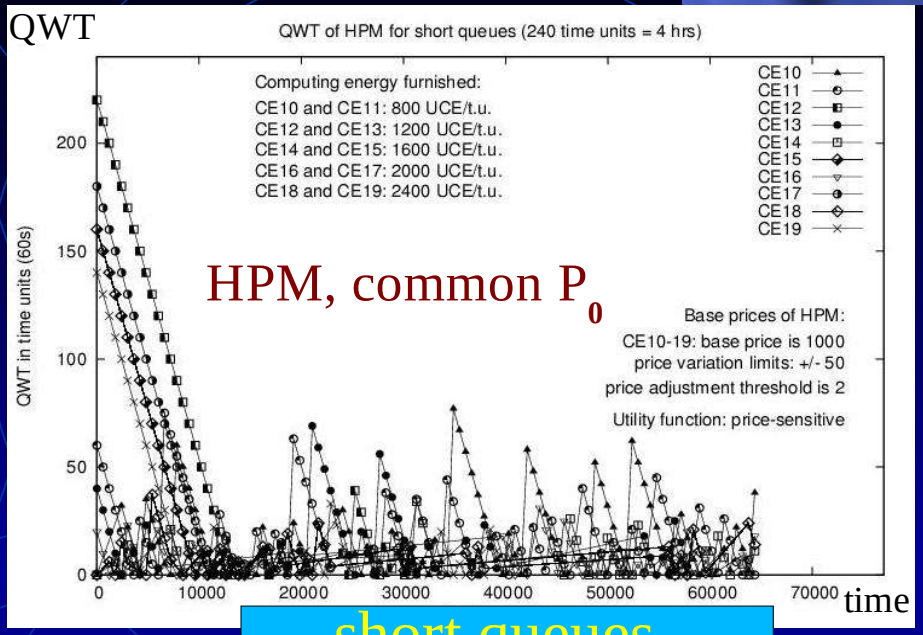
Importance of overlapping price domains

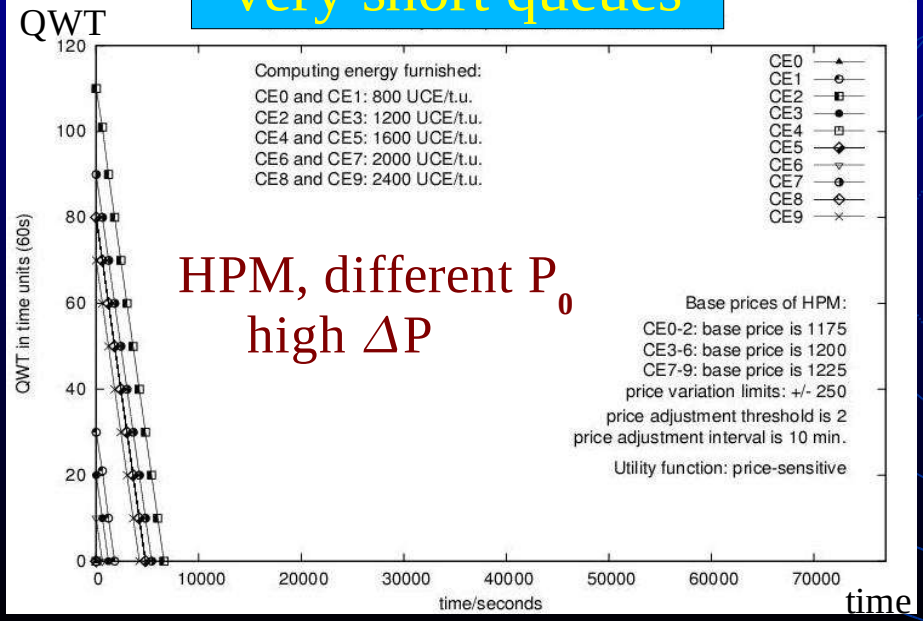
- if using price-sensitive brokering an effective load balancing requires widely overlapping price domains:
  - *Sim. run 2*: common base price of P_0 = 1000 GC/UCE for all CEs; price variation limit ΔP of 50 GC/UCE
  - *Sim. run 3*: base prices P_0 between 775 and 1225 GC/UCE (the same base prices as in sim. run 1); high price variation limit ΔP of 250 GC/UCE (=> price domains are widely overlapping!)
- non-overlapping price domains lead to "subgrids" with eventually different levels of workload (which may also be a desirable setting, where users pay more for immediate processing, e.g. for interactive jobs)

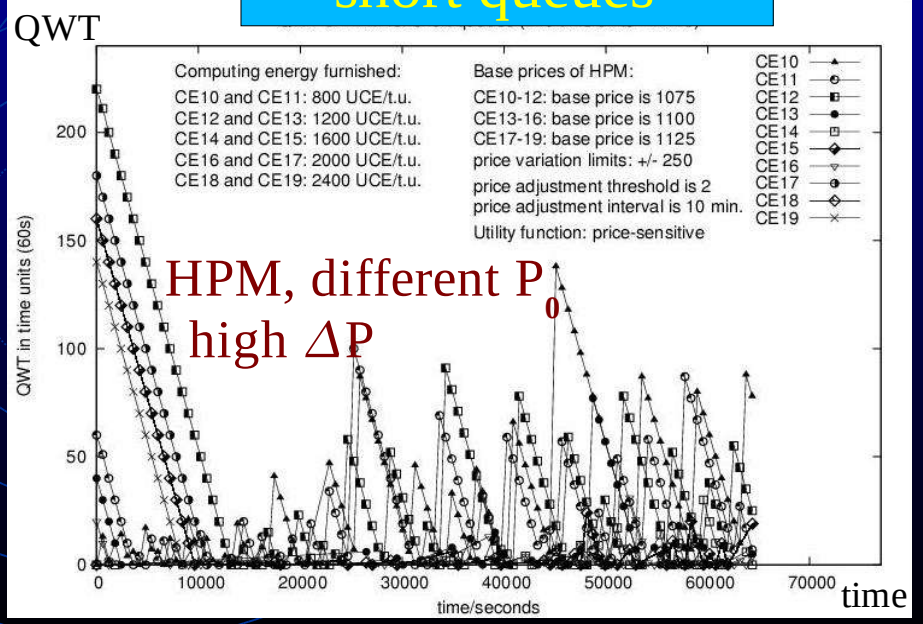
piro@to.infn.it

Results (7): QWT, very short & short queues

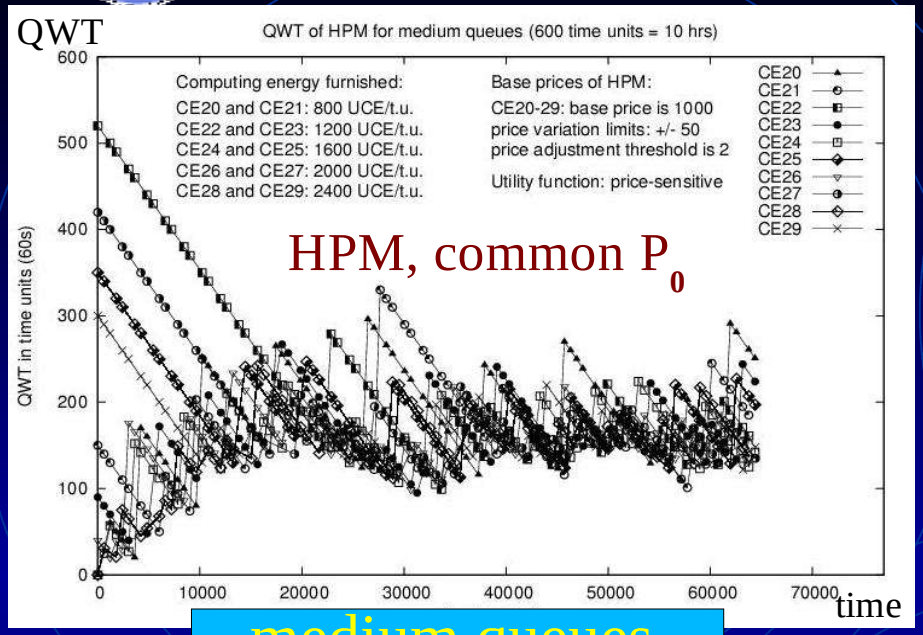




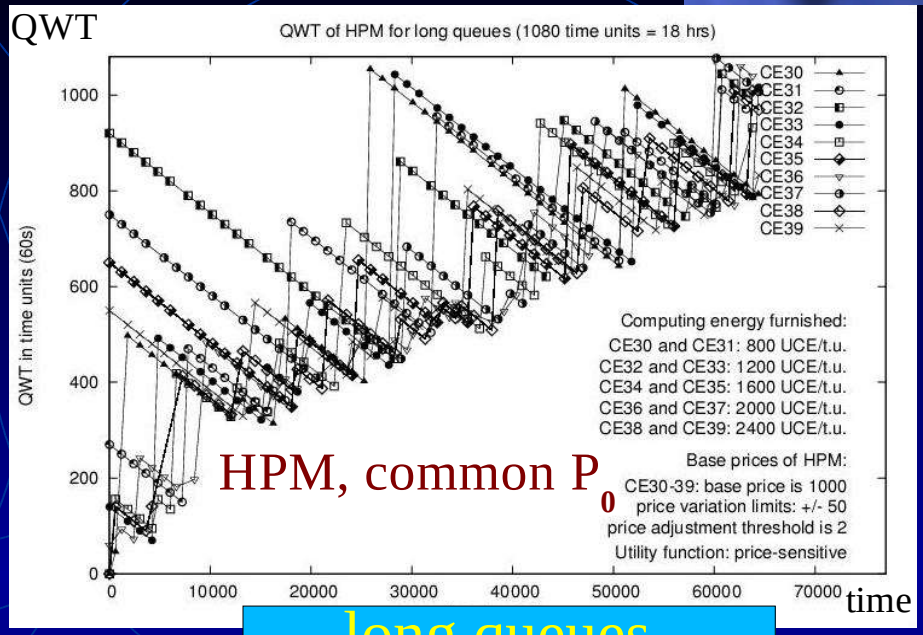




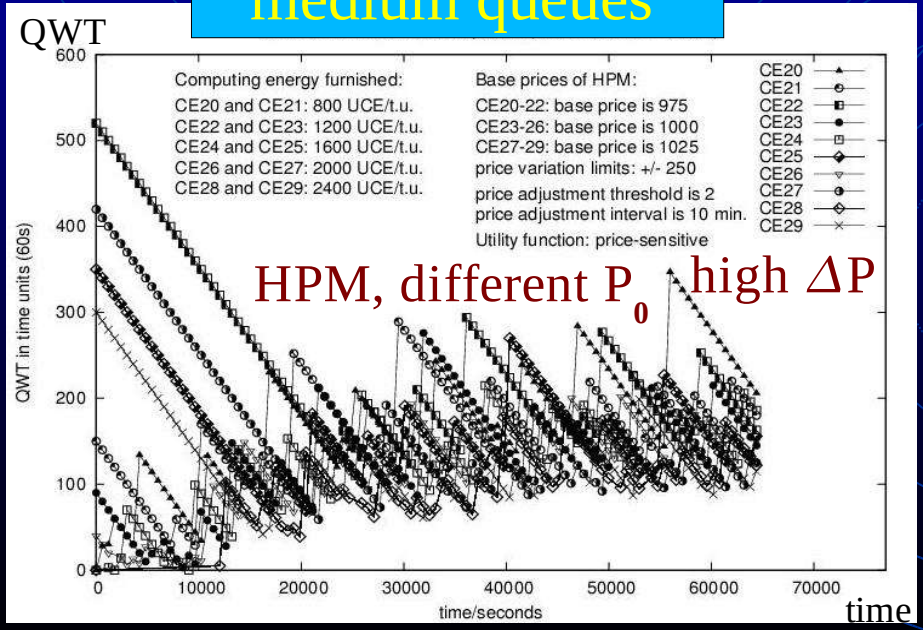
Results (8): QWT, medium & long queues

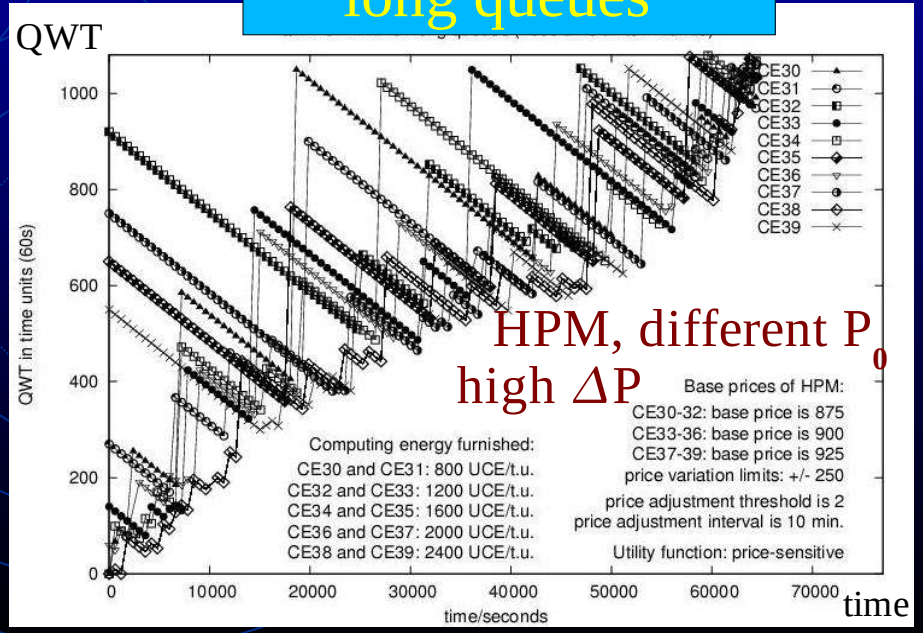


medium queues

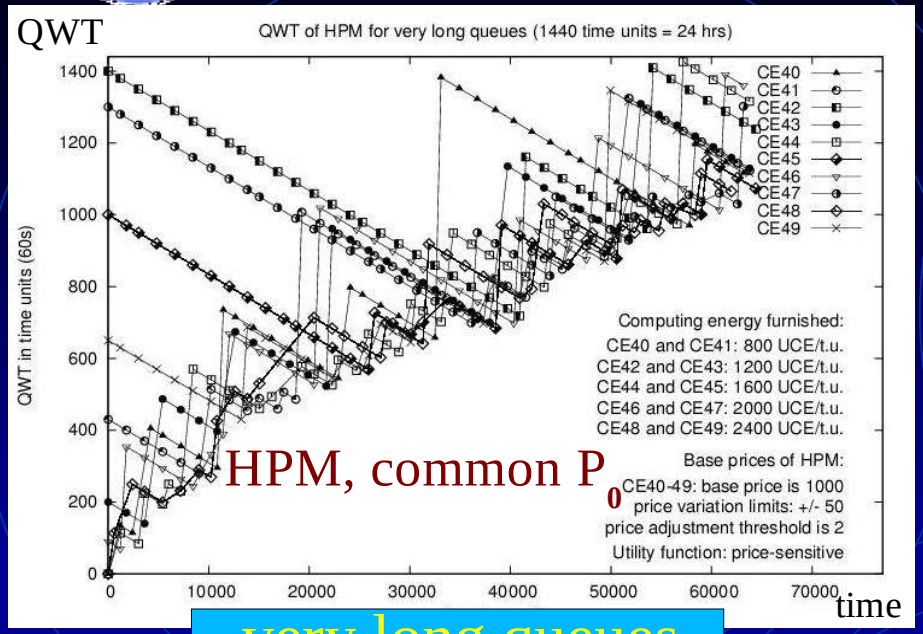


long queues

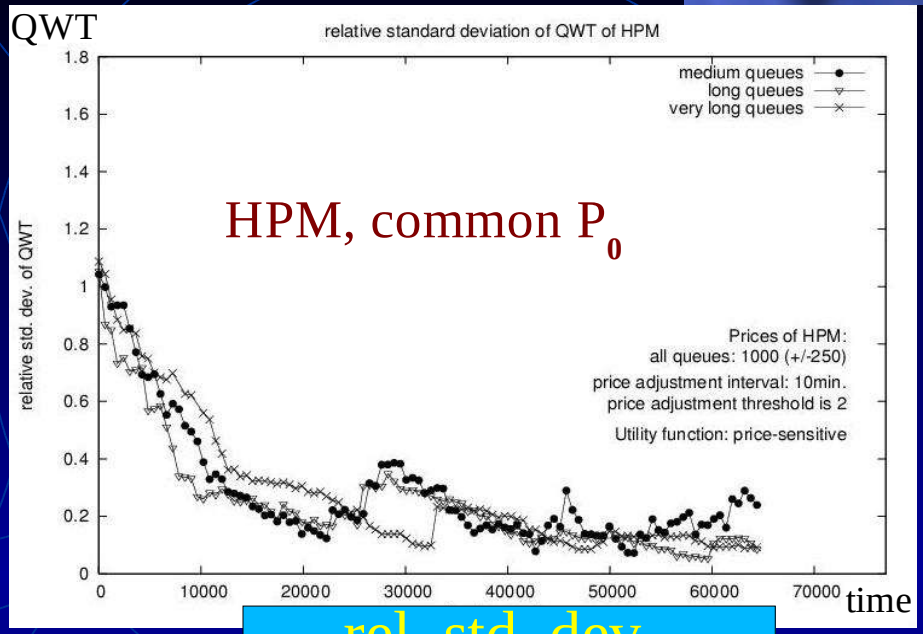




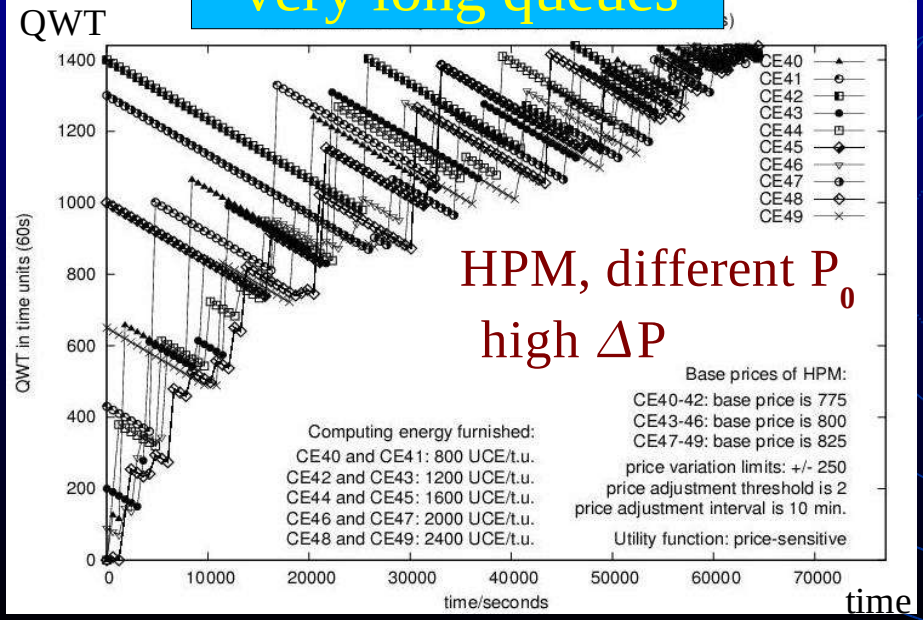
Results (9): QWT, very long q. & rel. std. dev.

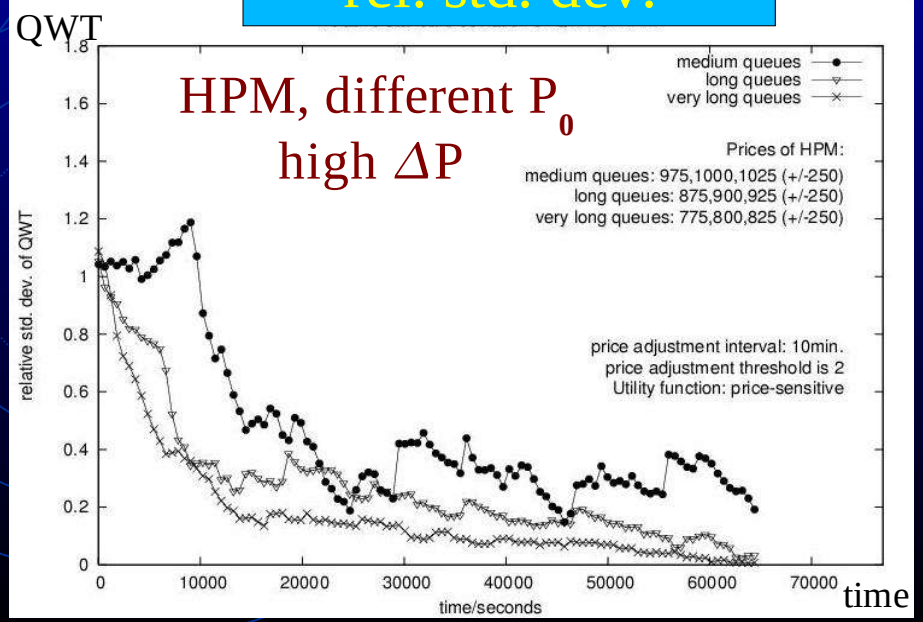


very long queues



rel. std. dev.





Conclusions

- DGAS-Sim allows to study the impact of different brokering strategies and pricing algorithms on load balancing in computational grids before deploying them on testbeds or production grids.
- most previous simulations of computational economies for grid computing focused on reaching a market equilibrium by computing an equilibrium price valid for all transactions, without considering the workload on the single resources.
- the Hybrid Pricing Model – provided an appropriate configuration – approximates the "best case" (completion time-based brokering), additionally offering the advantages of an economic approach:
  - possibility to reach market equilibrium (requested computational energy equals furnished computational energy) by incentivating users to delay less urgent jobs to periods with lower congestion.
  - an economic framework may include data management, data storage, networking and priority (limited budgets) issues.

piro@to.infn.it