Contribution ID: **51** Type: **poster**

Predicting Bounds on Queuing Delay in the EGEE grid

Wednesday 9 May 2007 17:30 (20 minutes)

Describe the scientific/technical community and the scientific/technical activity using (planning to use) the EGEE infrastructure. A high-level description is needed (neither a detailed specialist report nor a list of references).

Predicting the performance of schedulers is a notoriously difficult task [1]. As a consequence, grid users might be tempted to work around the standard grid middleware by designing specific strategies, which would be counterproductive if generally adopted. On the other hand, Machine Learning has been successfully applied to performance prediction in distributed and shared environments [2,3]. This paper reports on experiments on predicting the basic parameters of scheduling in the EGEE framework.

Report on the experience (or the proposed activity). It would be very important to mention key services which are essential for the success of your activity on the EGEE infrastructure.

We carried out preliminary statistical analysis (including summary statistics, density estimation, and time series analysis) on scheduler logs of a site of the EGEE grid (the LAL node). We show that the experimental arrival process and service times are extremely far from simple standard models (the classical M/M/N Kendall queue model with Poissonian arrival times and exponential service time), and might in fact exhibit long-range correlation and periodic behaviour. The failure of linear autoregression suggests that non-linear methods are more appropriate in the time series analysis of the expected queuing delay. We are currently investigating such methods (neural networks, gaussian processes and hidden Markov models), which can be able to take into account both inter-arrival time and load.

With a forward look to future evolution, discuss the issues you have encountered (or that you expect) in using the EGEE infrastructure. Wherever possible, point out the experience limitations (both in terms of existing services or missing functionality)

The major pitfall in analyses similar to our approach is the possible lack of representativity of the data. Further research in this direction could greatly profit from an easier access to the existing monitoring data (beyond isolated experiments). Furthermore, easier access would also reduce the associated cost of developing analysis software.

Describe the added value of the Grid for the scientific/technical activity you (plan to) do on the Grid. This should include the scale of the activity and of the potential user community and the relevance for other scientific or business applications

The expected running time (RT) of jobs and expected queuing delay (QD) are important inputs for grid global schedulers. Within gLite, QD is dynamically published by the Computing Elements into the grid information system, which is in turn queried by the scheduling agents called the brokers. At this time, little is known about the accuracy of the prediction of QD. In ordinary production, gLite uses the published QD for minimizing the expected job turnaround time, and errors in this prediction impact grid utilization. gLite also considers all jobs being equivalent, so it is difficult (without reconfiguring the site schedulers) to raise the priority of certain classes of jobs in situations such as social emergency, important events for a scientific community, or software prototyping. To overcome these problems, reinforcement learning has been proposed as a solution for time-constrained scheduling by coupling efficient prediction of QD and scheduling decisions.

Authors: Dr KÉGL, Balázs (CNRS); Prof. GERMAIN-RENAUD, cécile (CNRS LRI); Mr PEREZ, julien (LRI)

Presenters: Dr KÉGL, Balázs (CNRS); Mr PEREZ, julien (LRI)

Session Classification: Poster and Demo Session