
Contribution ID: **89** Type: **Poster**

# Statistical analysis of the workload and modeling of the time a job spends at different states in the EGEE environment

## 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).

In this study we analyze the workload of LCG/EGEE production Grid and the time the job spends at different states. The probabilistic models obtained for the job submission process and the delay components introduced at different stages is important for modeling production Grids such as the EGEE.

## 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.

Our results indicate that if we consider the LCG/EGEE Grid as the level of our observation, jobs are submitted without any specific weekly or daily patterns. The job interarrivals match very well with a rounded exponential distribution. We define and model four delay components that comprise the overall job processing. We observe that the total time of a job in the EGEE Grid is dominated by the CE's Register and Queuing time (D3) and the WN's Execution time (D4), while the other two components, namely, the time spent in the Pending, Submitted and Waiting states (D1), and the time in the Ready state (D2), are small.
We also evaluate the efficiency of the EGEE Grid and (indirectly) of the employed super-scheduling algorithm by comparing the total delay experienced by a job with that of a hypothetical ideal super-cluster. We found that we would have similar performance if we submitted the same workload to a super-cluster having 34% of the total average number of CPUs participating in EGEE

## 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

We have considered the overall LCG/EGEE infrastructure as a single entity and observed the general properties of job submission and execution, by collecting and analyzing the daily reports of the Real Time Monitor tool. Based on the EGEE job flow diagram we distinguish four delay components of the job processing, each corresponding to the time spent at different states in the EGEE environment. We analyze and model each delay component separately and we also model the job arrival process. The existence of good probabilistic models is important for the improved understanding of the Grid Computing concept, the prediction of their performance, the design of the middleware they use, and the evaluation of new scheduling and quality of service policies. We also examine the efficiency of the EGEE environment and (indirectly) of the employed super-scheduling algorithm.

**Author:** Mr CHRISTODOULOPOULOS, Konstantinos (University of Patras)

**Co-author:** Prof. VARVARIGOS, Emmanuel (University of Patras)

**Presenter:** Prof. VARVARIGOS, Emmanuel (University of Patras)

**Track Classification:** Demo and Poster session