



Contribution ID: 64

Type: **Oral**

Modeling Grid Job Time Properties

Wednesday, 14 April 2010 11:20 (20 minutes)

Large-scale analysis of grid job data, consisting of more than 28 million jobs gathered during 20 months from all major EGEE Resource Brokers, provides interesting results regarding the properties of job time parameters. The main parts in a job's life cycle are observed to have different distributions, having also different origins that cause these behaviors. While 'match time' parameter corresponds better to power law distribution, the total job length is well modeled with log-normal distribution. Here we attempt to fit the distributions and give explanation through generative models.

Detailed analysis

We can recognize three main parts in the life cycle of a job: time from registration to Resource Broker dispatching job to a suitable Computing Element (match time); time the job waits in CE's queue (wait time); and time it was actually being executed (run time). We can also observe the total lifetime of the job, the sum of the three, as a separate parameter, exploring also the stake of the parts in the total length. When analyzing distributions of these parameters, we note the straight line signature in log-log scale diagram, particularly for the 'match time' parameter, pointing to power law. However, the 'total time' parameter can be fitted more accurately with log-normal distribution, which comes as a consequence of the generative model considering the dependencies between consecutive jobs. Multiplicative processes, where in every step the size of the event (here, job length) grows or shrinks according to a random variable multiplier, can be applied here to give the explanation for the possible log-normal distribution.

Conclusions and Future Work

Close relations between power-law and log-normal distributions have already been noted in the literature. Very small variations in generative models are shown to decide between the two distributions of event sizes. As in the case of job length parameters, other generative models can be tested and applied, possibly bridging the gap between the two with the double Pareto, or double Pareto log-normal distributions.

Impact

The results of the large scale analysis of job length parameters give us insight into global behavior of the grid network. Power laws and log-normal distributions are often associated to natural processes and are related to emergent behavior of complex systems. In this case, understanding the distributions of time parameters can be used in network simulation, optimization, scheduling and self-management. Understanding how jobs are correlated between themselves and what kind of behavior it causes on global level is valuable information in many different contexts, proving it usable even for predicting future job parameters. Besides, more analysis is performed regarding the correlation between different time parameters of the same job (like efficiency, for example, which is defined as run time over total time), bearing useful results as well.

Keywords

Job length, distribution, run time, generative models, efficiency

URL for further information

www.grid-observatory.org

Primary author: Mr ILIJASIC, Lovro (University of Turin)

Co-author: Prof. SAITTA, Lorenza (University of Eastern Piedmont)

Presenter: Mr ILIJASIC, Lovro (University of Turin)

Session Classification: Computer Science

Track Classification: Scientific results obtained using distributed computing technologies