A Study of Grid Applications: Scheduling Perspective

Aleksandar Lazarević, Dr. Lionel Sacks

ACSE Group
Dept. of Electronic and Electrical Engineering
University College London
Outline

• Introduction to Grid Computing
• Study of Grid Applications
• Implications for Grid Scheduling
Grid Computing

- Utility platform
- Computational Service

- Range of Hardware

- Range of Uses/Applications
Mission Statement

• Study Grid Usage/Applications
  – Explore job & resource utilisation statistics
  – Patterns of user behaviour and workflows
  – Correlation with historical data

• Explore Resource Management implications
  – On higher level: planning, provisioning and SLA
  – On lower level: admission control and scheduling
UCL Grid Cluster

- 6 months of job accounting data from UCL’s Central Computing Cluster
- 25 eScience projects, ~ 50,000 jobs
- Collect meta-data on job submitter, submit time and node, scheduling delay, real CPU time, **wallclock execution time** etc.
Overall Job Execution Times
Group Execution Times
Group Job Distribution

Group Job Count Distribution
- matsim [85%]
- chemccs [4%]
- ocoir [4%]
- geogrs [7%]

Group Total Job Execution Time Distribution
- matsim [7%]
- chemccs [26%]
- geogrs [18%]
- ocoir [49%]
Cumulative Execution Time Distributions

Empirical CDF: all jobs

Wallclock Execution Time log[s] vs. Probability of Observation

Empirical CDF: geneog
Empirical CDF: chemcpp
Empirical CDF: ecceler
Execution Time Autocorrelation

Wallclock Execution Times Autocorrelation Function

Wallclock Exe Times Partial Autocorrelation Function
Scheduling Implications

- Could we anticipate job duration & resource requirements trends?
- Could we predict job duration of specific jobs in the queue?
- Could we offer an intuitive “deadline” scheduler with low administration cost?
- Could users live with probabilistic guarantees?
Conclusions

- Observed emergence of workflow patterns and their correlation with job meta-data
- Heterogeneous application set, size of jobs related to the resource size and expected performance
- High levels of autocorrelation could make statistical modelling feasible
Q & A

- a.lazarevic@ee.ucl.ac.uk
- www.ee.ucl.ac.uk/~alazarev/
- Self-Organising Grid Resource Management (SO-GRM) Project: www.ee.ucl.ac.uk/acse/so-grm/index.htm