

Measuring and Monitoring Grid Resource Utilisation



London Communications Symposium
2004

Aleksandar Lazarević
Supervisor: Dr. Lionel Sacks
University College London



Outline

1. Grid Management Problems
2. Project Overview
3. Monitoring Framework
4. Grid Load Simulation
5. Further Scheduler Research

Grids are persistent environments that enable software applications to integrate instruments, displays, computational and information resources that are managed by diverse organizations in widespread locations.

Is It Here To Stay?



“The emerging field of eScience should transform this kind of work. It’s significant that the UK is the first country to develop a national eScience grid, which intends to make access to computing power, scientific data repositories and experimental facilities as easy as the web makes access to information.”

Tony Blair, October 2002



the globus toolkit™

www.globustoolkit.org



SO-GRM Overview

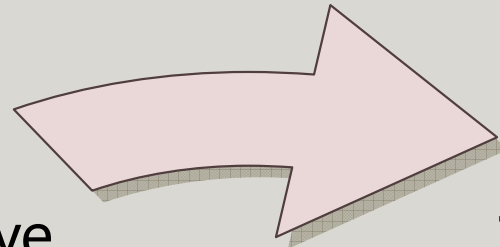
- Self-Organising Grid Resource Management
 - 3 year UK e-Science EPSRC sponsored project in collaboration with BT @ Adastral Park
 - Base research in autonomous Grid management

- Developed components
 - SORD: Self-Organising Resource Discovery
 - I3: Autonomous Security Monitoring & Enforcement

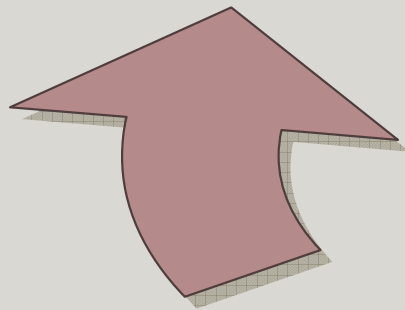
- In Development
 - SLA Management
 - Monitoring & Scheduling

Methodology

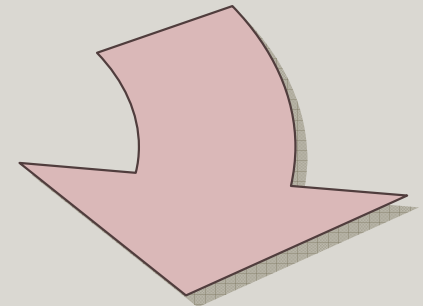
Granular and Adaptive
Resource Monitoring



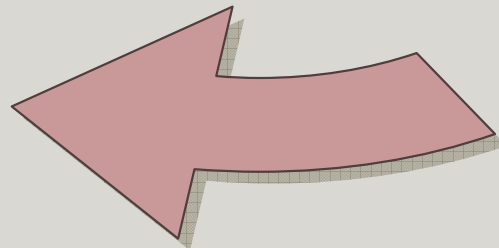
Statistical Analysis:
Job Arrival & Service rates
Resource Utilisation
Execution Time
Social Factors



Schedule Performance
Monitoring Feedback
Confidence Level
Improvement

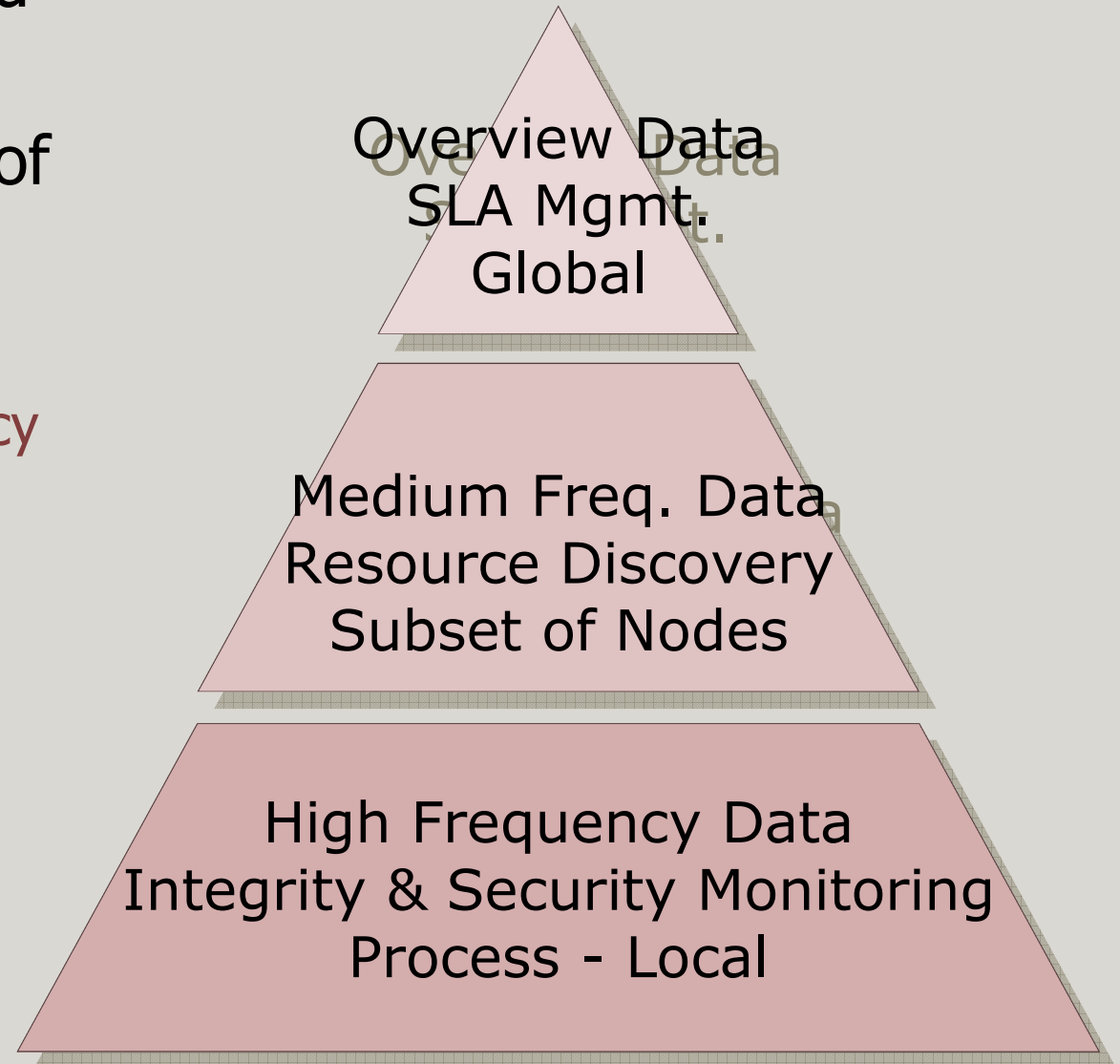


Predictive, Probabilistic
User-Orientated
Deadline or Economy
Based Scheduler

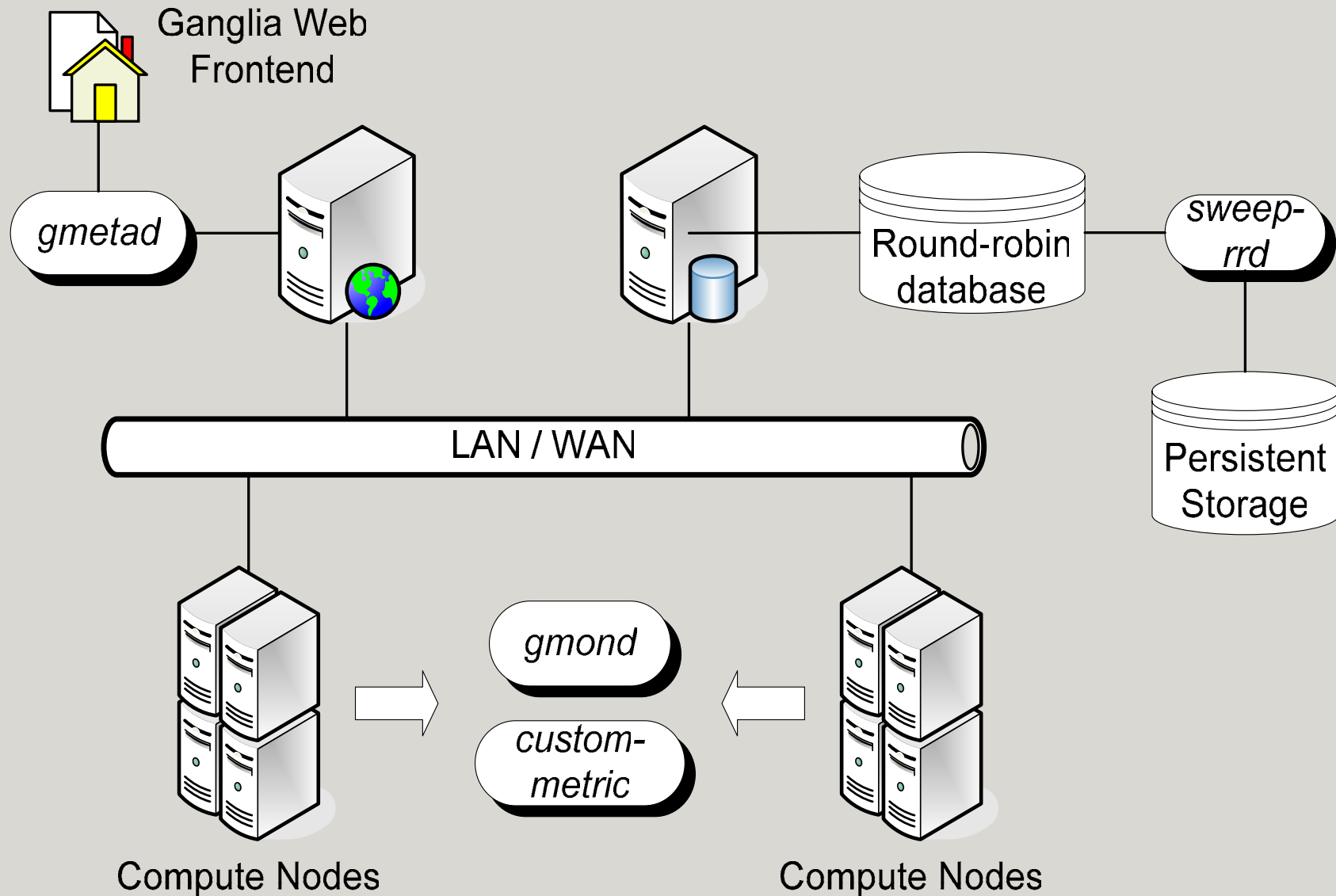


Monitoring and Measurement

- Separation of volatile and non-volatile metrics
- Hierarchical distribution of monitoring data
- One data base adapted:
 - By measurement frequency
 - By Scope
 - By Consumer
- Data communication method adaptable:
 - Distributed \Rightarrow effective
 - Centralised \Rightarrow reliable

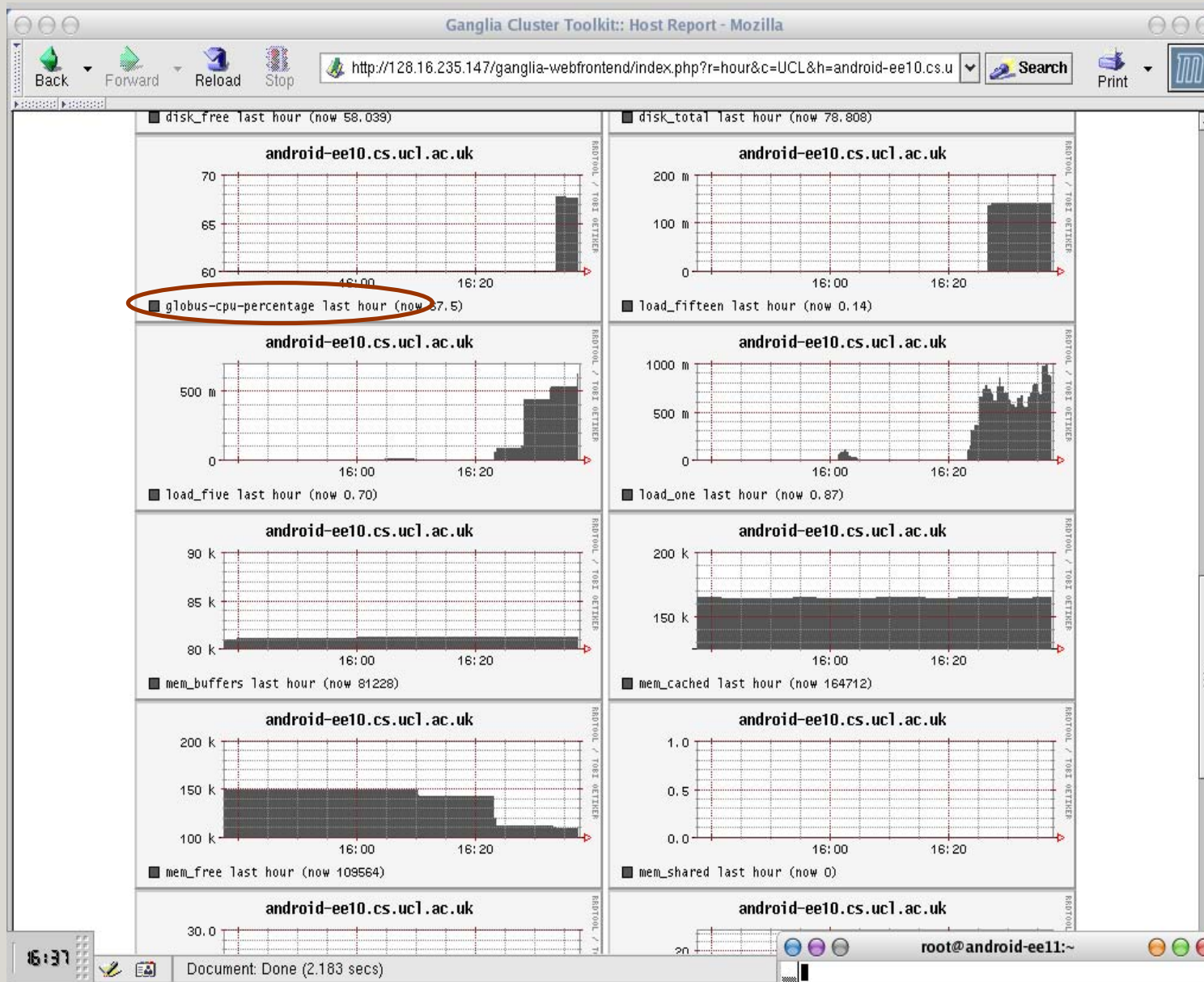


Monitoring Functional Diagram



Ganglia Monitoring

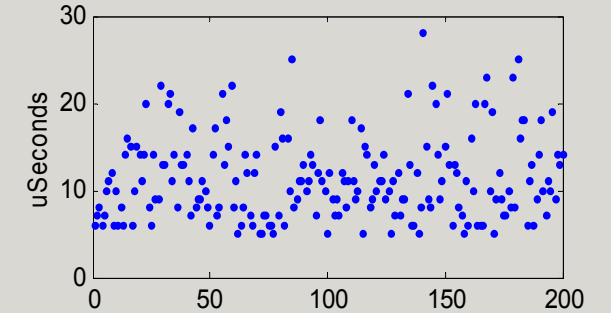
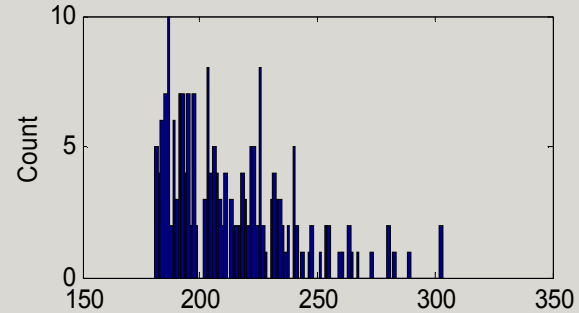
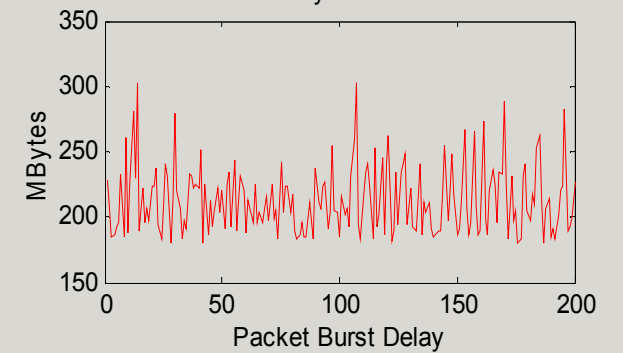
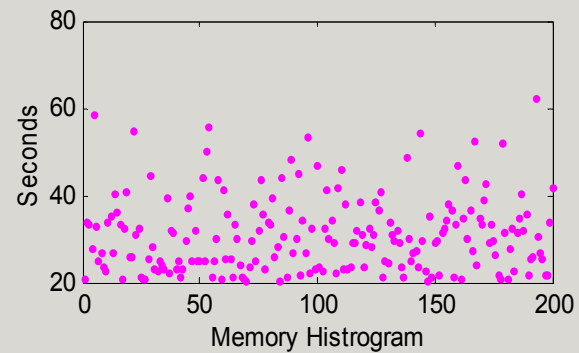
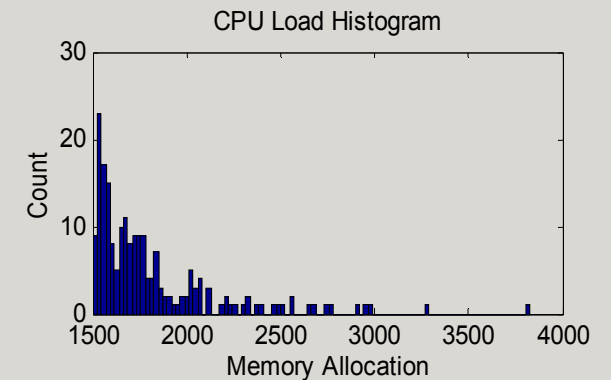
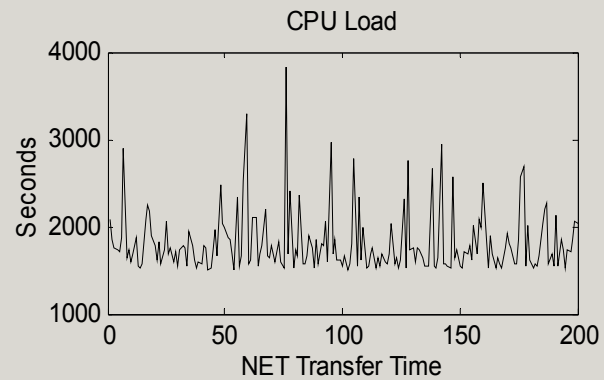
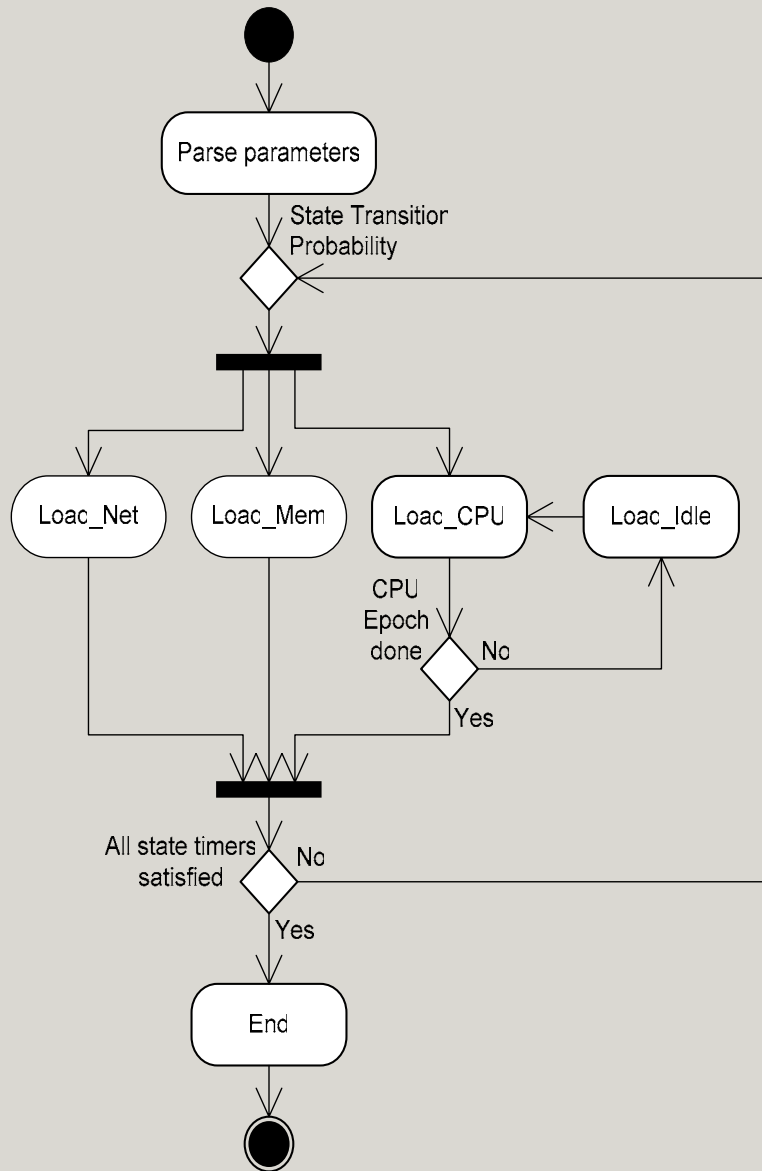
Ganglia Node Monitoring



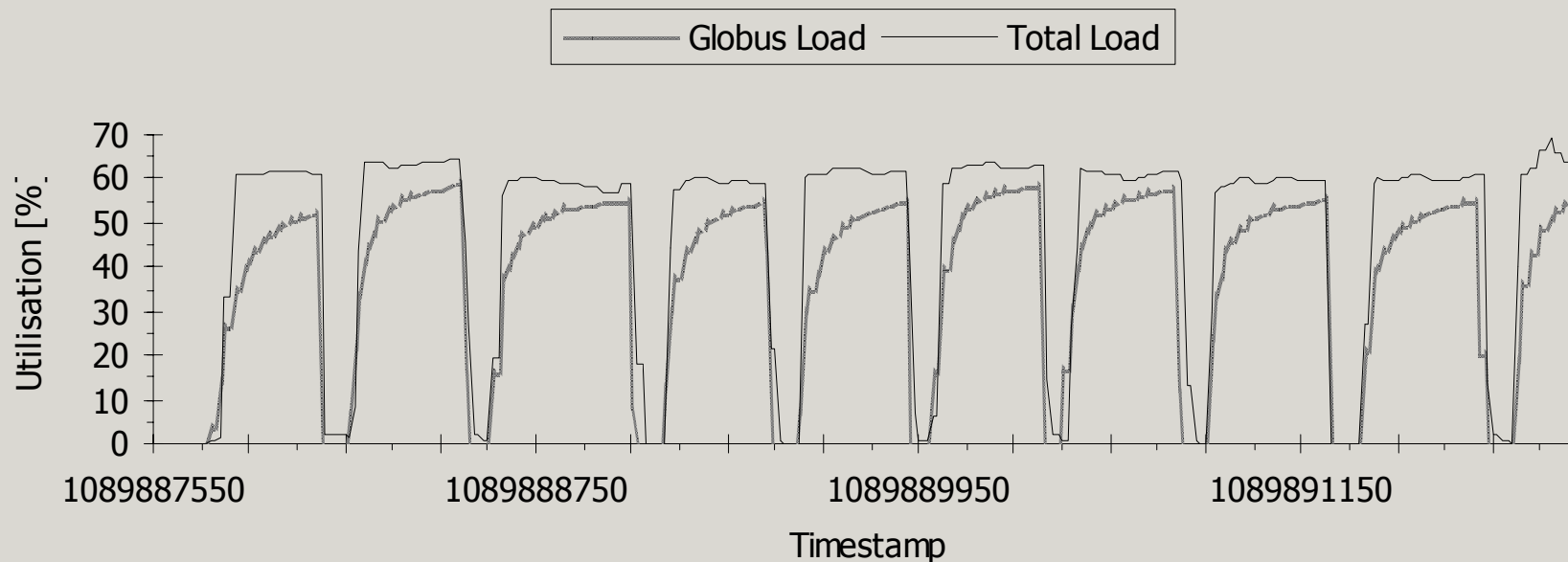
Grid Load Simulation

- Scheduler testing problems:
 - Few simulation tools: SimGrid and GridSim
 - Little knowledge of Grid application statistics
 - Problems modelling dynamic and heterogeneous Grid resources
- Testing in self-organising and adaptive context:
 - Presented computational load must be variable and dynamic
 - Test runs must be controllable and repeated
 - Overall job arrival distribution must be controllable

GridLoader



GridLoader – Test Results



Adaptive Scheduler Research

- Lightweight, user transparent approach rules out
 - Application re-compilation or source code changes
 - Resource utilisation prototyping or instrumentation
 - Resource intensive prediction
- Develop a fast one-step-ahead prediction based on statistics of previous runs
- Exploit time and space locality of jobs and target machines
- Gradually improve confidence levels

Conclusions

- Grid will require intelligent and autonomous managing components
- SO-GRM developed components:
 - SORD – small-world based distributed resource discovery
 - I3 – distributed intrusion detection
 - GridLoader – parameterised emulation of Grid applications load
 - Monitoring framework – highly granular, adaptive, integrated with other management components
- Further research: probabilistic Grid scheduler

Q & A

- UCL Research Computing:
www.grid.ucl.ac.uk
- More info on SO-GRM:
www.ee.ucl.ac.uk/~alazarev/