Adaptive Grid Scheduling and Resource Management

Adrian Li Mow Ching, Ioannis Iliabotis, Aleksandar Lazarević,
Tope Olukemi, Dr. Ognjen Prnjat, Dr. Lionel Sacks
University College London

SO - GRM

"Self-organized Grid Resource Management" is an EPSRC funded e-Science project in cooperation with BT ExacT.

SO-GRM draws its motivation from the lack of new-breed, adaptive, autonomous and robust resource management and scheduling frameworks for heterogeneous computing environments.

SO-GRM researches several key areas of Grid computing:

1. High-level policy based VO management and Service Level Agreement (SLA) negotiation
2. Light-weight, resilient and self-organizing resource monitoring and discovery methods
3. Adaptable and dynamic job scheduling based on previously observed performance data and future predictions
4. Run-time process auditing and security monitoring

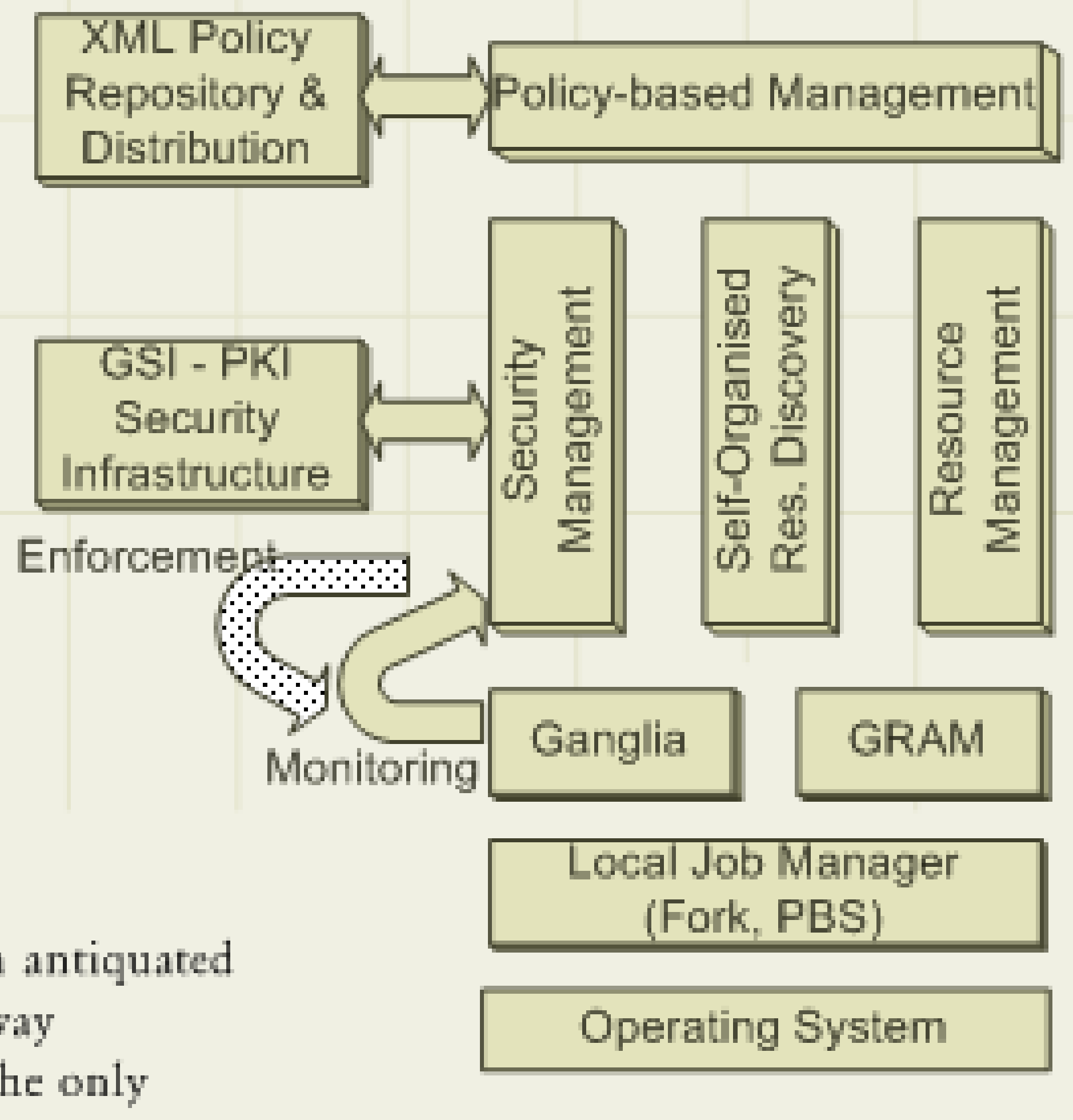
Workload Distribution

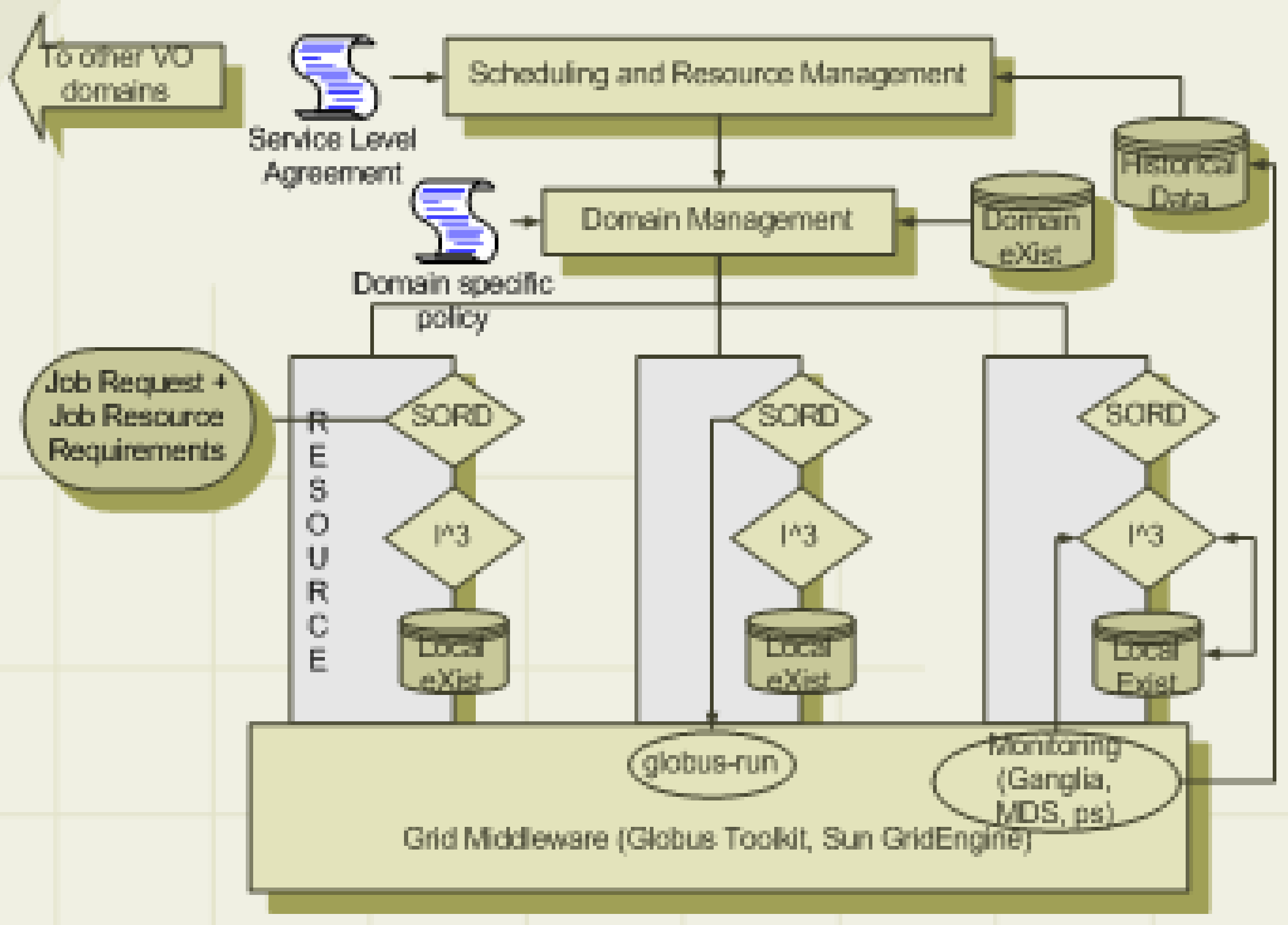
The key component of the SO-GRM project is the workload distribution protocol and its ability to place jobs presented to the system on an appropriate node as efficiently as possible. Rather than maintaining a central data base of cluster utilisation, the nodes in the system learn about sub-groups of nodes and about other sub-groups. By constructing a small-worlds network through the system, available resources can be located quickly through short interrogation phases. Each node adapts to the success or failure of historical interrogations and provides feedback to the shape of the 'social' network used to allocate work. In practice there may be several overlaid and interweaved small-world networks, each specialising in different classes of service, providing QoS based resource allocation without significant extra computational load.

Framework

Components of SO-GRM framework are built using widely accepted open-source technologies:

1. XML Policies stored in open-source databases such as eXist
2. Integration of Globus Security Infrastructure (GSI) and using Public Key Infrastructure (PKI) based on X.509 certificates
3. Scalable monitoring system based on Ganglia Monitoring Toolkit exchanging XML messages via broadcast/ multicast/unicast
4. Support for various local job managers overlaid with Globus GRAM
5. Wide operating system support


XML Policy Repository & Distribution
Policy-based Management
GSI - PKI Security Infrastructure
Security Management
Self-Organised Res. Discovery
Resource Management
Enforcement
Monitoring
Ganglia
GRAM
Local Job Manager (Fork, PBS)
Operating System



Components

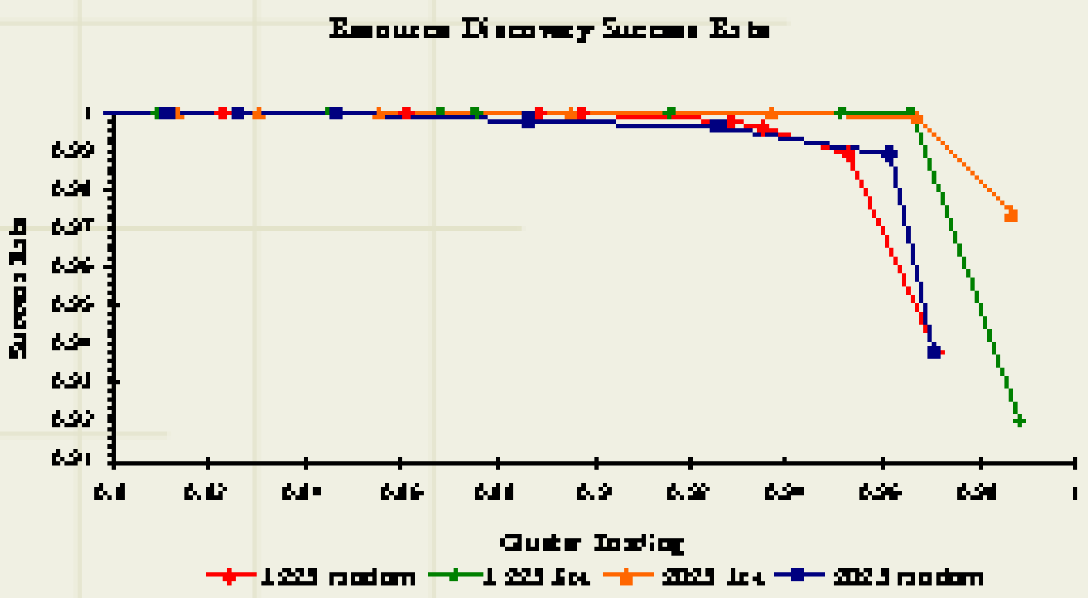
Many current scheduling/queuing approaches are based on antiquated master-worker model. The performance of the master gateway determines the service rates for incoming jobs and, being the only route for job submission, renders it a single point of failure.

SO-GRM approach is to develop a distributed system of independent compute nodes, all able to accept job requests and choose whether to execute or forward them to a more appropriate node. The levels of acceptable loading, quality of service guarantees and SLA enforcement can all be set through policy-enabled management. Multiple job entry points and self-organizing information dissemination based on small worlds topology should significantly increase resilience.

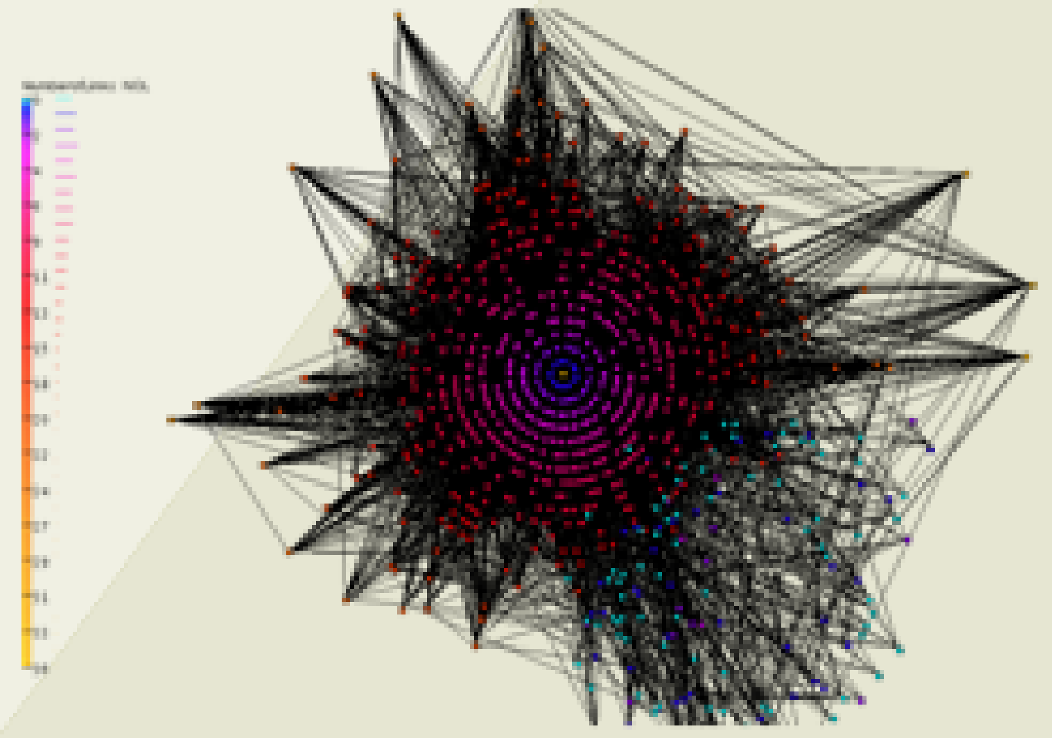
Resource Discovery

Resource discovery has been implemented through fully decentralised Java-based agents (SORD) and an open XML protocol. Nodes in the topology are initially connected to a number of their nearest neighbours and few distant nodes, representing the shortcuts through the network. As requests are received this topology evolves into a small-worlds network, and nodes begin to differentiate in frequency of fulfilling certain type of resources or classes of service queries.

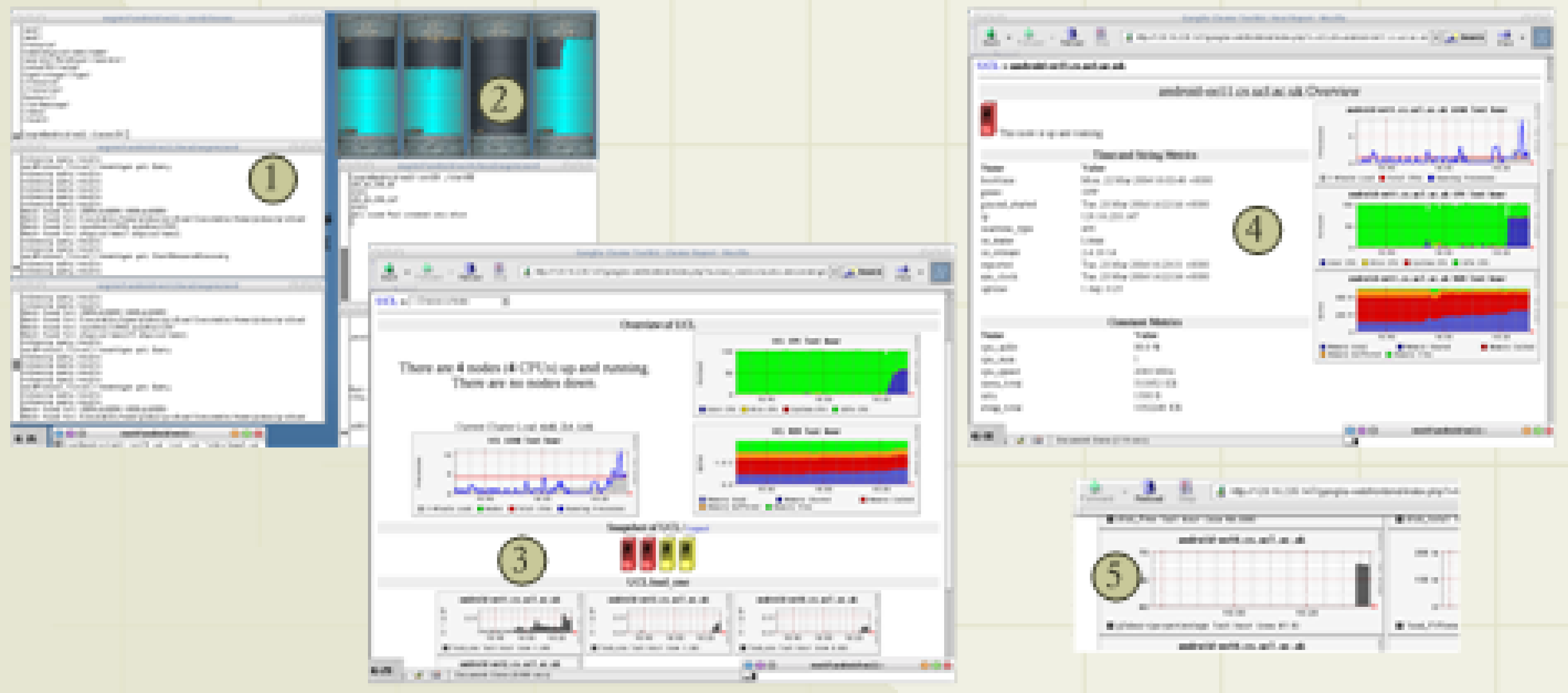
The protocol scales very well, with model simulations and live testing showing a high discovery success rate even for extreme overall cluster utilisation percentage.



Evolved Network Topology



19500 requests, 1225 nodes, 6 near + 1 far neighbour



Demo

Components integration was demonstrated on a multi-domain test-bed compromising computational resources in BT@Adastral Park and UCL. Thorough monitoring and information capture was essential during test runs, this task being further complicated by the probabilistic nature of SO-GRM's resource discovery and allocation algorithm.

1. Proper functioning of resource discovery is followed through debug messages
2. Machines are being assigned jobs. Load is only one of parameters taken into consideration, thus an already loaded machine may be assigned another job before an idle node.
3. Ganglia monitoring system showing a snapshot of UCL Grid
4. In-depth information on single execution node.
5. Additional information providers have been developed to report CPU utilization of Globus submitted jobs.