IEEE DS ONLINE EXCLUSIVE CONTENT

Works in Progress

Cluster Computing and Grid 2005 Works in Progress

The first in a two-part series of works-in-progress articles from the Cluster Computing and Grid 2005 conference held in Cardiff, UK. For more information on each of these projects, contact the authors.

Grid-Oriented Storage

Frank Wang • **Cambridge-Cranfield High Performance Computing Facility**
Na Helian • **London Metropolitan University**
Sining Wu, Yuhui Deng, and Ke Zhou • **Cranfield University**
Yike Guo • **Imperial College**
Steve Thompson, Ian Johnson, Dave Milward, and Robert Maddock • **Xyratex**

Because the grid will eventually become an open market, grid service negotiation must not only focus on computing jobs but also data storage. At the Cambridge-Cranfield High Performance Computing Facility Centre for Grid Computing, we're working on a project sponsored by the UK's Engineering and Physical Sciences Research Council and the UK Department of Trade and Industry called Grid-Oriented Storage. GOS aims to support advanced data bank services and data reservoirs so that multiple computers and end users can share data on the grid.

Our GOS appliance (see figure 1) fits the thin-server categorization and is the first of its kind to implement grid computing at the device or hardware level. A GOS appliance is a disk array that connects directly to the grid using a Globus XIO interface. To applications running on the grid, GOS looks like a large-capacity hard disk. To a client on the grid, it resembles an ordinary file server. We've installed an Apache server on the GOS server, which allows HTTPS-based communication between the GOS server and a client via a Web browser.

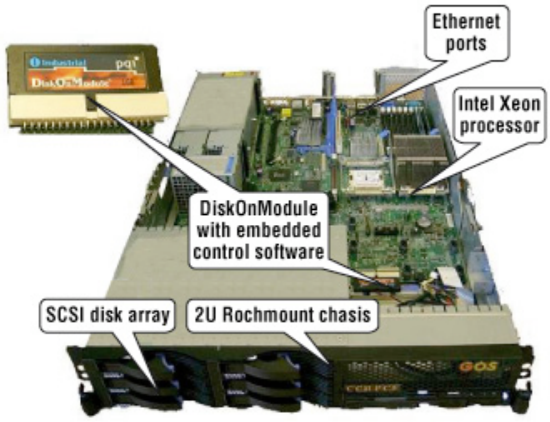

Figure 1. The Grid-Oriented Storage appliance.

We based the GOS security model on the Grid Security Infrastructure (GSI). The GOS requests and receives a certificate from the Certificate Authority. The GOS must also be registered in a virtual organization's Lightweight Directory Access Protocol service.

GOS provides scalable storage bandwidth without the cost of servers that are used primarily for transferring data from peripheral networks to client networks or grids. GridFTP, pipelined network transfer and disk access, secure interfaces via GSI and block-level security, and the increased availability of DiskOnModule motivate and enable this new architecture.

Tests of our prototype system show that we can integrate GOS appliances into the Open Grid Service Architecture. GOSFTP improves data transfer performance by 20 to 40 percent—a gain we attribute to the single-purpose intent of the GOS products and their corresponding I/O-optimized designs. Furthermore, we've shown scalable bandwidth for GOS-specialized file systems. Using a parallel data mining application, GOS appliances deliver a linear scaling per client-drive pair.

A GOS appliance functions in a GOS cluster in a peer-to-peer relationship, where each GOS appliance manages its own internal storage space. Clustered GOS appliances permit a common, aggregate presentation of the data stored on all the participating GOS appliances. When a user accesses data from the GOS cluster, a GOS appliance either fulfills the request directly (if the file resides in its storage) or requests the data from the GOS appliance managing it.

GOS appliance will offer a cost-effective alternative to server-tethered storage products on the grid, and significant R&D efforts are under way to expand GOS capabilities. GOS promises to take file server functionality down to the device level, while P2P GOS clusters have the potential of elevating the role of GOS into an all-encompassing data bank service solution on the grid.

For more information on this project, contact Frank Wang at frankwang@ieee.org.

Frank Wang is a professor and chair of grid computing and e-science and is the director of the Centre for Grid Computing, all at the Cambridge-Cranfield High Performance Computing Facilities. Contact him at frankwang@ieee.org.
Na Helian is a lecturer of data mining at the London Metropolitan University. Contact him at n.helian@londonmet.ac.uk.
Sining Wu is a research officer for the Centre for Grid Computing at Cranfield University. Contact him at s.wu@cranfield.ac.uk.
Yuhui Deng is a research officer for the Centre for Grid Computing at Cranfield University. Contact him at y.deng@cranfield.ac.uk.
Ke Zhou is a visiting professor at the Centre for Grid Computing at Cranfield University. Contact him at k.zhou@cranfield.ac.uk.
Yike Guo, Technical Director, London e-Science Centre, Parallel Computing Centre, Imperial College. Contact him at yg@doc.imperial.ac.uk.
Steve Thompson is CTO of Xyratex. Contact him at steve_thompson@xyratex.com.
Ian Johnson is a chief scientist at Xyratex. Contact him at ian_johnson@xyratex.com.
Dave Milward is a project manager at Xyratex. Contact him at dave_milward@xyratex.com.
Robert Maddock works at Xyratex. Contact him at bob_maddock@xyratex.com.

A Dynamic Estimation Scheme for Fault-Free Scheduling in Grid Systems

Benjamin Khoo • National University of Singapore

The dynamism of grid computational environments, which resources can join and leave in an unpredictable fashion, allows for unprecedented growth in the Grid's capacity. At the same time, resource availability at any given moment is uncertain.

In work under way at the National University of Singapore, we've observed that most grids fall into one of two main categories:

- *Planned and negotiated*, in which the infrastructure is designed with resources committed to the grid.
- *Dynamic and voluntary*, in which the infrastructure operates in a peer-to-peer fashion with resources joining or leaving on no prearranged schedule.

This discrepancy makes it difficult to design robust and fault-tolerant scheduling algorithms for both resource categories. It also leads to difficulties in providing capabilities such as advanced reservations and quality-of-service assurances.

Figure 2 illustrates a resource lifecycle model to address this problem. It creates a series of time events and applies stochastic methods to the probability of each one occurring. Any scheduling algorithm can use this model to quantitatively estimate how many nodes would be online at a certain time. Algorithms can also use this model to determine the probability of a job completing its execution based on its required runtime and node allocation. This information helps allocate resources efficiently.

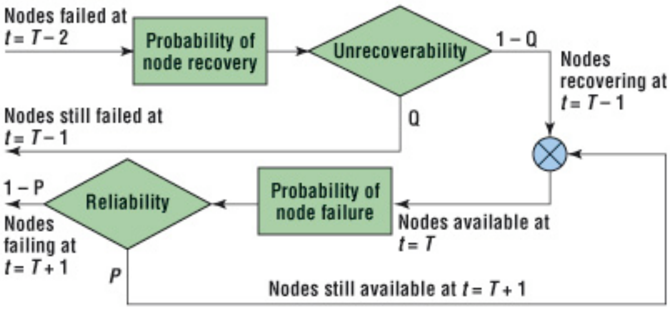

Figure 2. Resource lifecycle model describing the time window t that affects the number of nodes available at time *T*. *P* and *Q* describe the reliability and unrecoverability factors that modifies the basic MTTF and MTTR values.

Initial simulations based on this model have indicated prediction errors in the range of +/- 1 to +/- 2 CPUs in a Grid environment, when performed over a span 100 times longer then the mean time to failure (MTTF) values of the nodes.

Our next research objective is to incorporate this information in conventional scheduling algorithm heuristics over a homogeneous failure-prone environment and assess its effectiveness. We will then expand the work's scope to include large-scale heterogeneous distributed systems and grids.

For more information on this project, contact Benjamin Khoo at g0402607@nus.edu.sg.

Benjamin Khoo is a doctoral student at the National University of Singapore. Contact him at g0402607@nus.edu.sg.

Cite this article:
Frank Wang, Na Helian, Sining Wu, Yuhui Deng, Ke Zhou, Yike Guo, Steve Thompson, Ian Johnson, Dave Milward, and Robert Maddock, "Cluster Computing and Grid 2005 Works in Progress: Grid-Oriented Storage," *IEEE Distributed Systems Online*, vol. 6, no. 9, 2005.

Benjamin Khoo, "Cluster Computing and Grid 2005 Works in Progress: A Dynamic Estimation Scheme for Fault-Free Scheduling in Grid Systems," *IEEE Distributed Systems Online*, vol. 6, no. 9, 2005.

dsonline.computer.org

Send Link
Printer Friendly

IEEE Distributed Systems Online | ISSN: 1541-4922 | http://dsonline.computer.org