

# GridPP

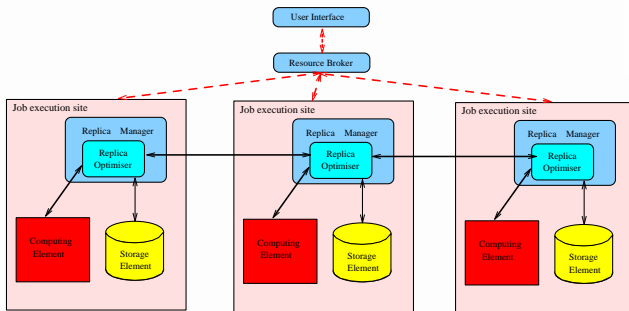
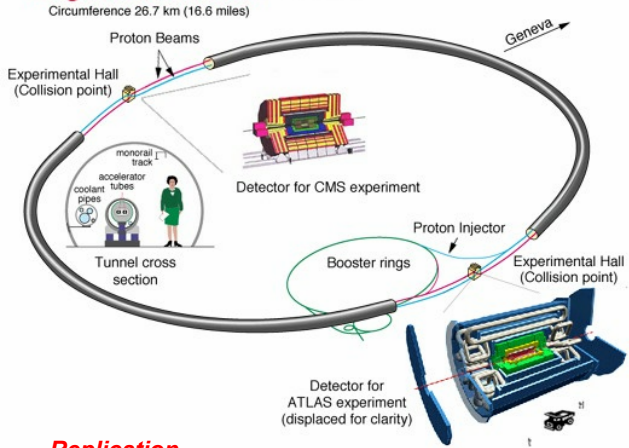
UK Particle Physics

[www.gridpp.ac.uk](http://www.gridpp.ac.uk)

## Dynamic Grid Optimisation

Intelligent middleware is key to the operation of large-scale data grids. Along with collaborators from the European DataGrid, we are working to construct data management middleware for the petabyte-scale UK Grid for Particle Physics (GridPP), which is currently preparing for analysis of data from the Large Hadron Collider at CERN, due to start production in 2007. Final middleware products must efficiently manage the use of Grid resources: network bandwidth, storage capacity and processing power of the Grid. Minimisation of network loading and maximisation of storage usage is possible by implementation of dynamic file replica management.

## Large Hadron Collider at CERN



European DataGrid Architecture

## Replication

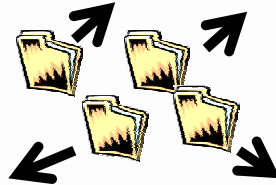
File replication is the process of creating copies of files at different Grid sites, and is known to improve performance. In the European DataGrid architecture, a replica manager service is present at every site, providing optimised file access via an internal replica optimiser. The role of the replica optimiser is twofold: it must provide the most efficient access to data for currently executing jobs, and through dynamic replica creation reduce data-access latencies for all users of the Grid.

## OptorSim

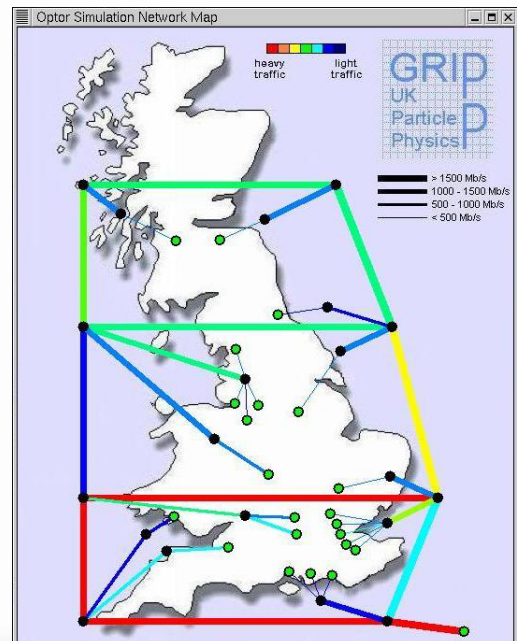
It is important to test possible replica optimisation algorithms before their deployment on the Grid, and this has led to the construction of a Grid simulator called OptorSim. OptorSim provides an artificial data grid infrastructure to allow testing of many different replication strategies. A developer using the software is able to describe site-to-site network connections, together with individual site resources.

The simulation aims to mimic a real Grid closely, including effects such as non-Grid traffic on the network. Furthermore, the simulation Provides a framework for a developer to describe site policies and job descriptions, where jobs are described by a list of associated files.

Most Grid simulators focus on job scheduling algorithms; OptorSim includes optimisation at both scheduling and job execution time, allowing the developer to mix and match strategies. When the simulation starts, jobs are distributed via a Resource Broker to Computing Elements within the Grid, using the scheduling algorithm. The Replica Manager at each site then finds the best files for the job, using the run-time optimisation algorithm to make the crucial decisions of whether or not to replicate the file locally and, if the local storage is full, which file(s) to delete in order to make space for the new replica.



OptorSim provides output in terms of several important metrics for the evaluation of replication strategies. These include the mean job execution time, computational power usage across the Grid, use of storage and network usage.



National  
e-Science  
Centre

e-Science  
eti

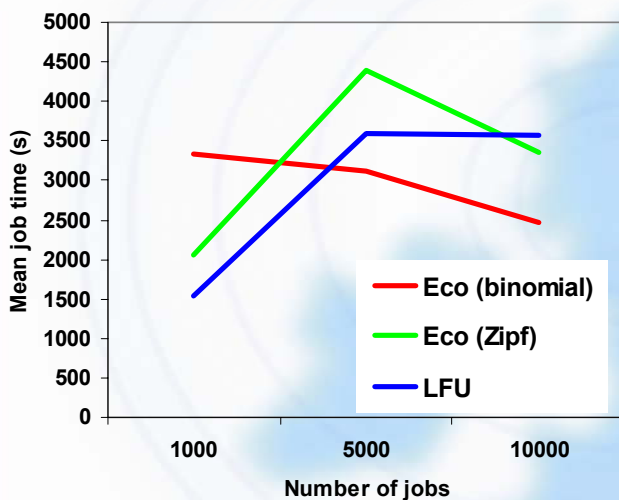
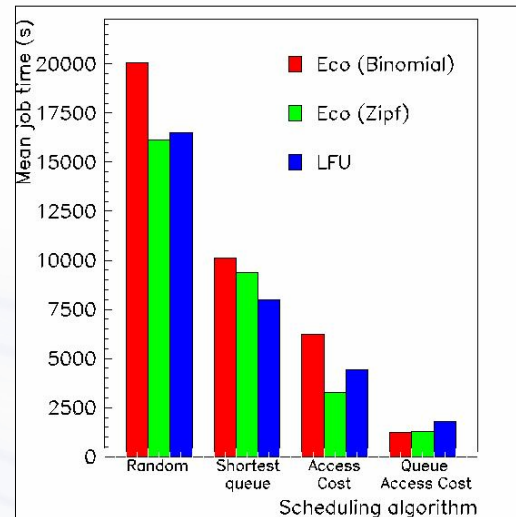
'Delivering e-science'

### Scheduling Optimisation Strategies

OptorSim offers 4 scheduling algorithms:

- *Random* - send the job to a random site.
- *Access Cost* - send the job to the site for which all the necessary files can be accessed most quickly.
- *Queue Size* - send the job to the site with the shortest queue.
- *Queue Access Cost* - send the job to the site for which all the necessary files for all the jobs in the queue – including yours – can be accessed most quickly.

The plot on the right shows the performance of these schedulers with the three run-time optimisation algorithms described below. *Queue Access Cost* is the best scheduler, giving a good balance between network usage and CPU load.



### Run-time Optimisation Strategies

We have used OptorSim to compare two novel economic models with a traditional Least Frequently Used (LFU) algorithm. The LFU model looks up all replicas in a catalogue before choosing the best, deleting the file which has been accessed least frequently in the recent past. The economic models use a Vickrey auction protocol to find the best replica, deleting the file of lowest value. The value of the file in the future is estimated using a prediction function based on the recent access history, and this is what differentiates the two models: the binomial economic model uses a binomial distribution to predict the future value of files, whereas the Zipf economic model uses a Zipf-life distribution. The sample plot on the left shows that, as the number of jobs on the Grid increases, the economic models begin to out-perform the LFU algorithm for the particular configuration under investigation.

### Other Parameters

The plot on the right shows the effect of including background, non-Grid traffic in the simulation, illustrating that this slows jobs down considerably. In this instance, realistic input from the UK-wide GridPP testbed was included in the simulation.

Different patterns of file access can also be simulated, reflecting the fact that different types of job may have different behaviours. Physics analysis jobs are likely to request files sequentially, whereas biomedical applications may follow other patterns. OptorSim includes sequential, Gaussian random walk and Zipf-distributed access patterns.

### Future Plans

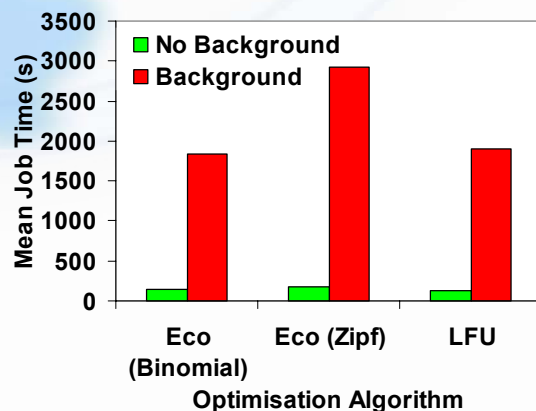
We plan to continue making OptorSim as realistic as possible, the next step being the inclusion of output job files. All this will help us find the best algorithms for replica optimisation, which can then be implemented on real Grids.

### Further Information

The EU DataGrid project: <http://www.eu-datagrid.org/>

GridPP: <http://www.gridpp.ac.uk>

OptorSim: <http://cern.ch/edg-wp2/optimization/optorsim.html>



### Further Reading

W.H. Bell, D. G. Cameron, L. Capozza, A. P. Millar, K. Stockinger, F. Zini *OptorSim – A Grid Simulator for Studying Dynamic Data Replication Strategies*, International Journal of High Performance Computing Applications, 17(4), 2003.

W.H. Bell, D. G. Cameron, R. Carvajal-Schiaffino, A. P. Millar, K. Stockinger, F. Zini *Evaluation of an Economy-Based File Replication Strategy for a Data Grid*, in International Workshop on Agent-based Cluster and Grid Computing at CCGrid 2003, May 2003.