



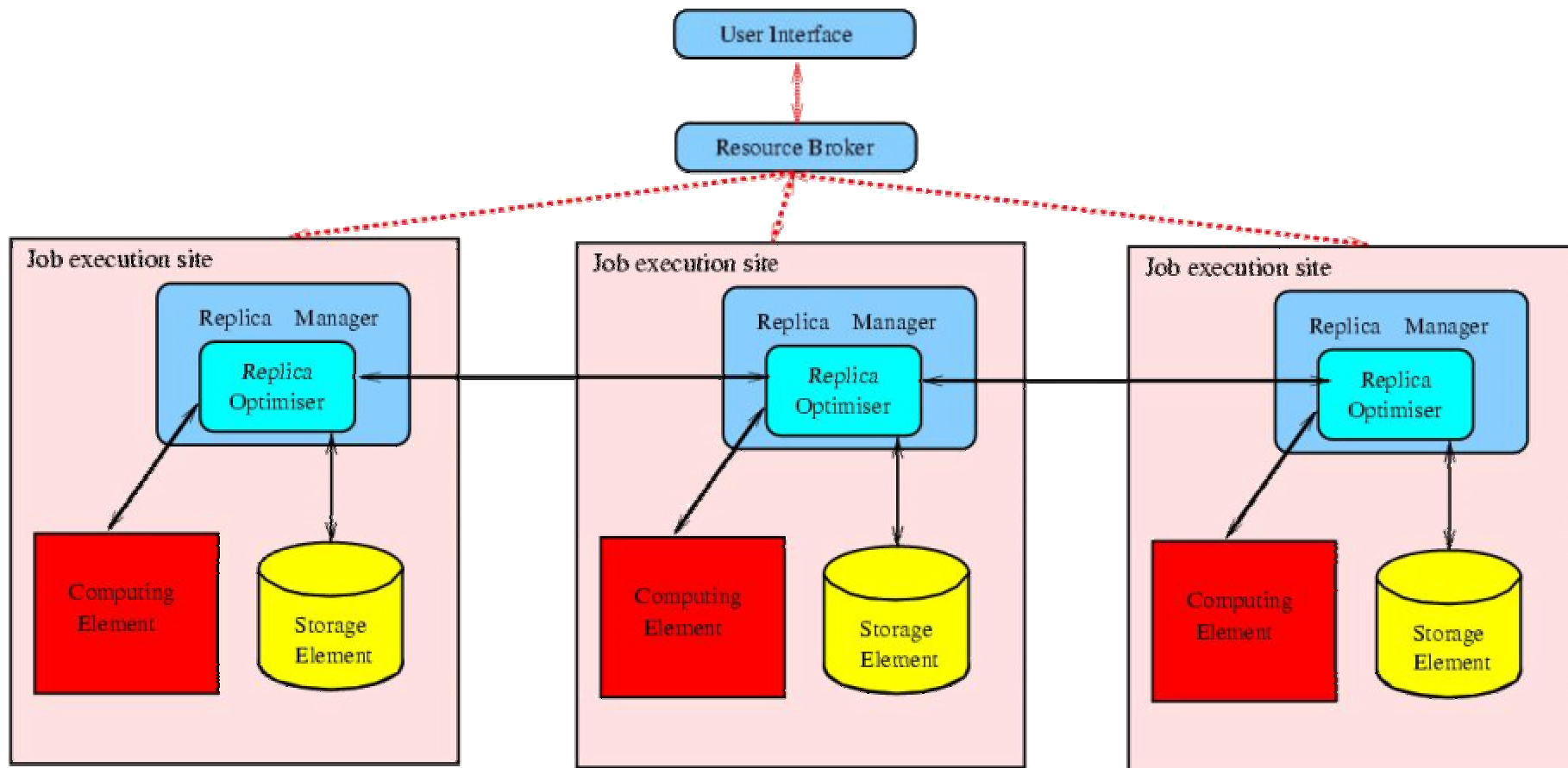
Simulation of Replica Optimisation Strategies for Data Grids

David Cameron
University of Glasgow





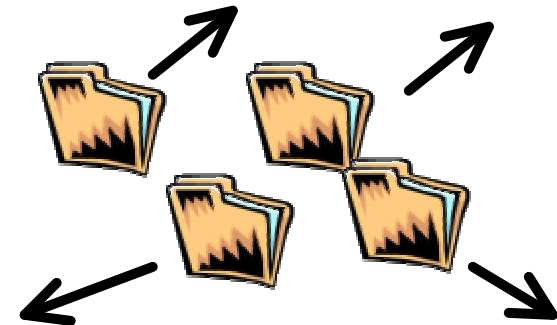
DataGrid Architecture





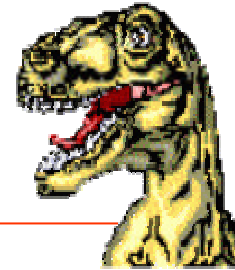
Grid Optimisation

- There are 3 stages in the lifetime of a job where optimisation occurs:
 - **Scheduling** - “find the best site to run my job”.
 - **Replica Selection** - “find the best replica for my current job” (short term optimisation)
 - **Dynamic Replica Optimisation** - “make sure replicas are in the best position for possible future jobs” (long term optimisation)

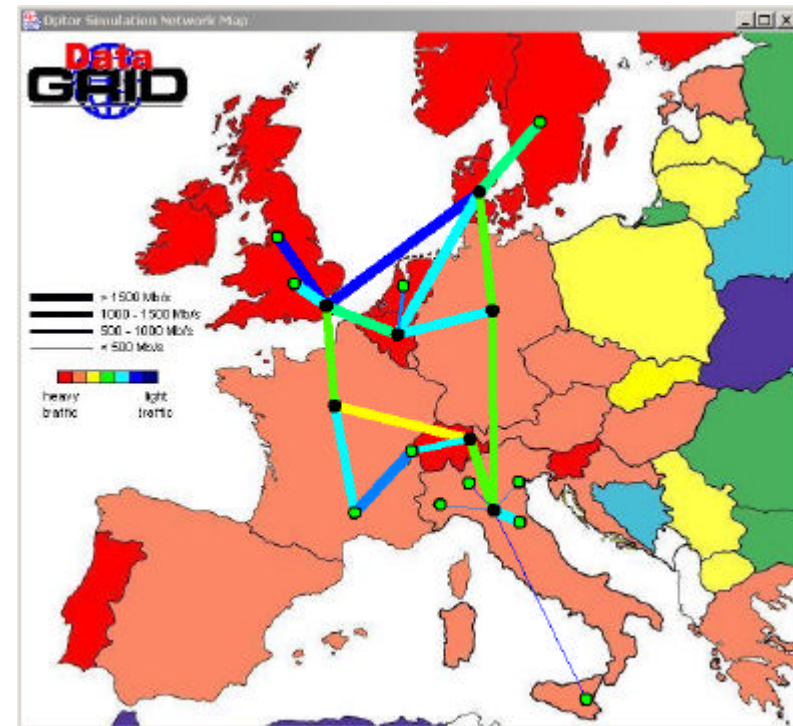




What is OptorSim?



- **OptorSim** is a **Grid simulator** written in Java to model the behaviour of certain **replica optimisation** algorithms
- It mimics a realistic Data Grid environment by **simulating** the execution of **experiments** that require distributed data
- It allows **testing** and **evaluation** of optimisation **algorithms** in various Grid scenarios



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



An Economic Model for Replica Optimisation

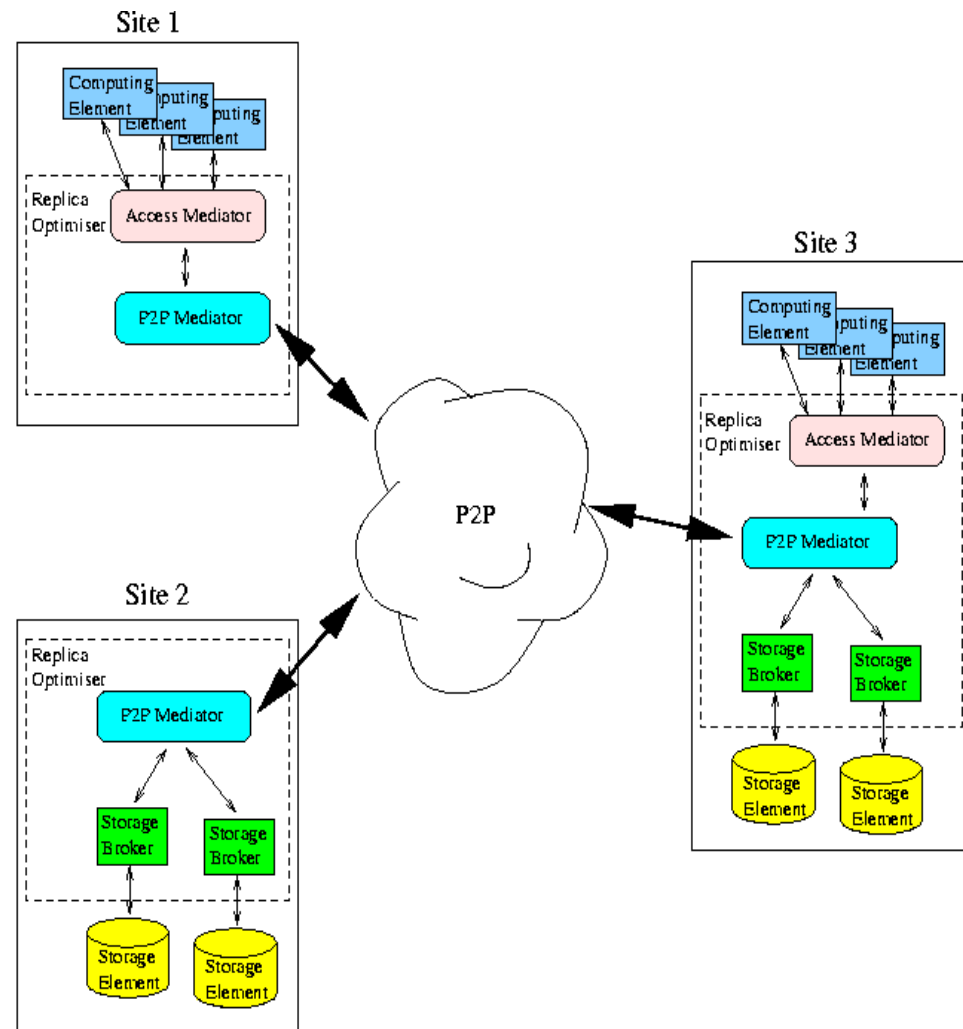
- Solves both **short-term** and **long-term** Replica Optimisation
- **Files** are **digital assets** which can be bought and sold for profit
 - **Computing Elements** (CEs) buy from **Storage Elements** (SEs) for running jobs
 - **SEs** “invest” in files to **sell** for profit to **CEs** and other **SEs**
- Both CEs and SEs interact via P2P network with intelligent independent optimisation agents which perform the reasoning using a **prediction function**
- Advantages: **distributed** and **dynamic**





P2P structure of Replica Optimiser

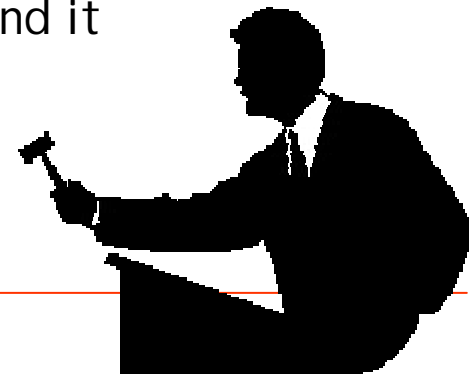
- **Access Mediator (AM)** - contacts other replica optimisers to locate the cheapest copies of files for the Computing Element.
- **Storage Broker (SB)** - manages files stored in storage element, trying to maximise profit for the finite amount of storage space available.
- **P2P Mediator (P2PM)** - establishes and maintains P2P communication between grid sites.





Auction Protocol for Replica Optimisation

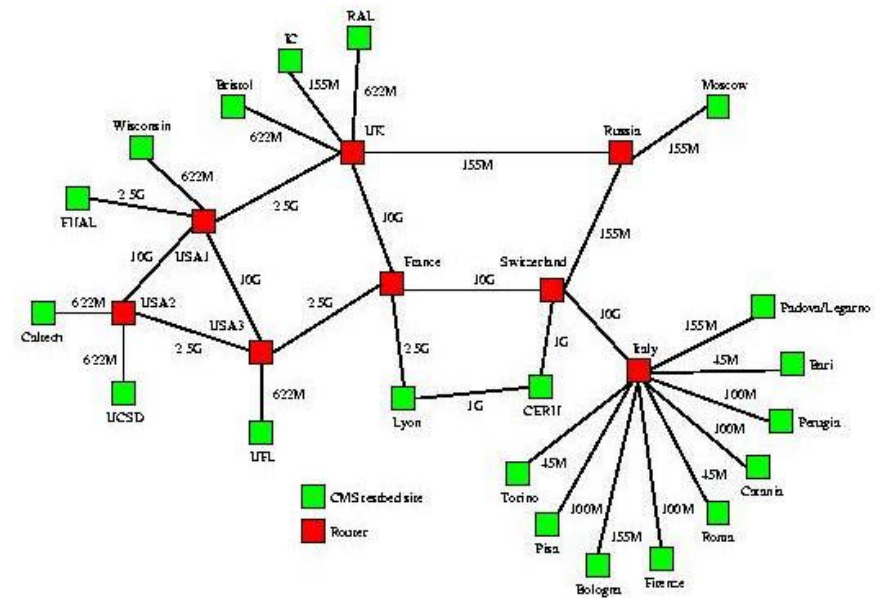
- ◆ We need a mechanism to **fix the price** of a file sold by a SB to an AM (or another SB) that **guarantees**:
 - **Low** price for purchaser
 - Trading **fairness**
 - Minimal messaging / **fast** as possible
- ◆ We use a **Vickrey auction** (sealed bid auction):
 - Every potential seller makes an offer (lower than or equal to the proposed maximum price)
 - The seller that made the lowest offer is chosen, and it is **paid** the **second-lowest offer**
 - Currently **bid** is based on **file transfer cost**



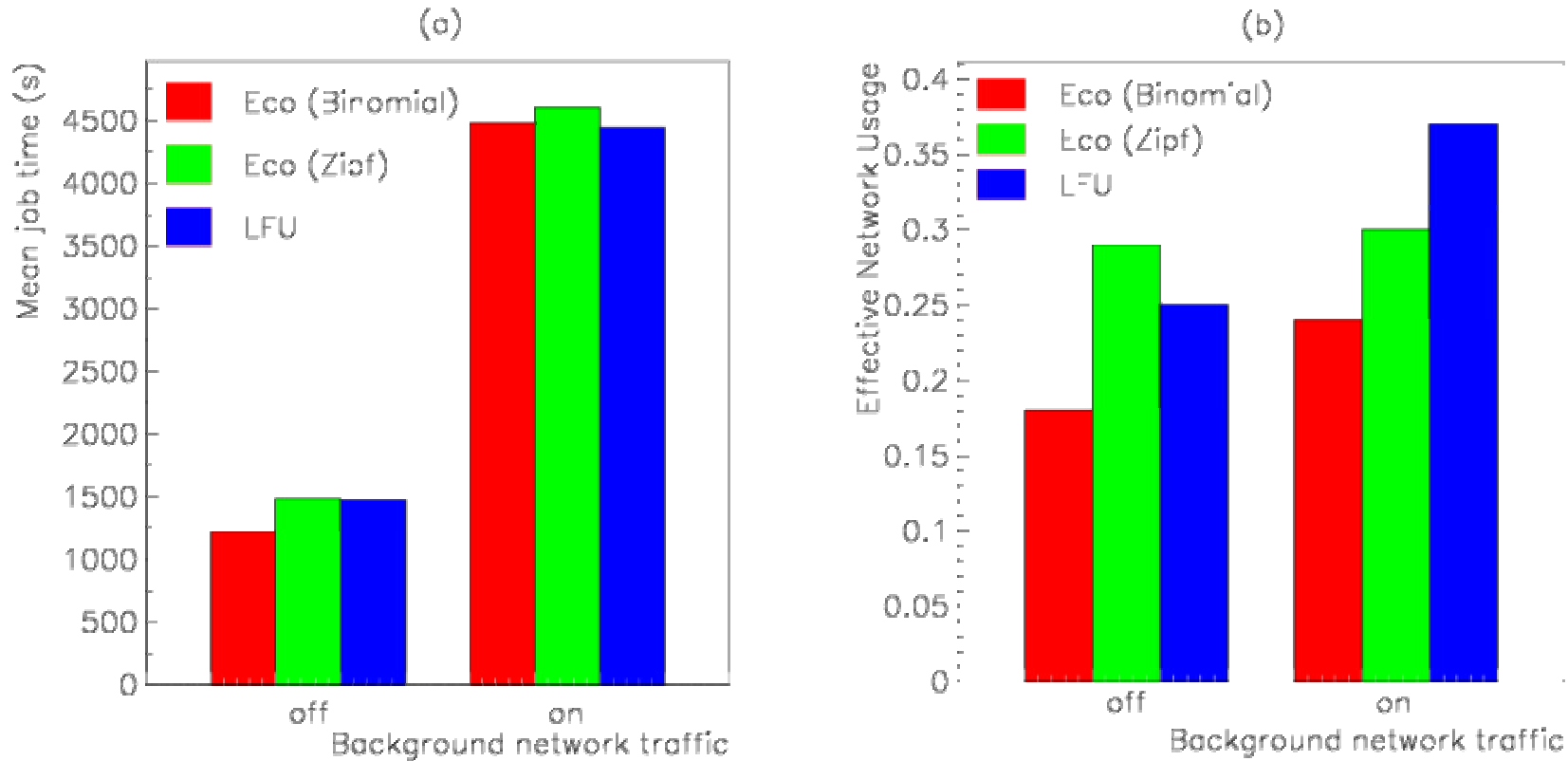


OptorSim Set-Up

- Use **CMS Spring 2002 Testbed**
 - 20 sites (Europe + US)
 - 6 countries
 - Take into account background network traffic
- Measure the time for the Grid to complete a number of **physics analysis jobs**.
 - Jobs based on real CDF analysis jobs
 - Total file size 97 GB
 - SEs sizes @ CERN and FNAL 100 GB, all other sites 50 GB
 - Initially all master copies are @ CERN and FNAL
- Compare two economic models and Least Frequently Used (LFU) algorithm.



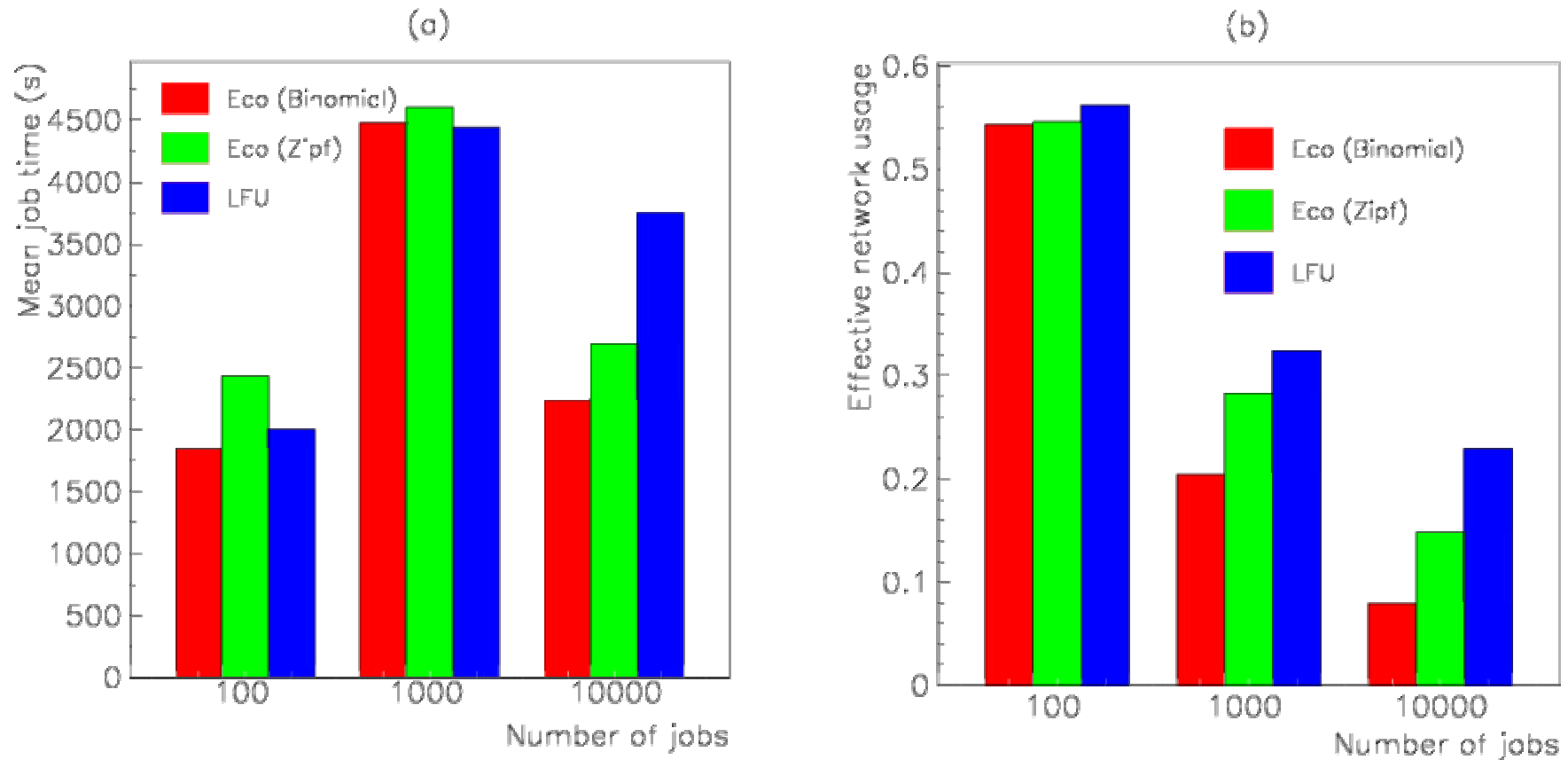
Effects of Network Traffic



Large **increase** of **mean job time** and moderate increase in network use when simulating background network traffic.



Mean Job Time & Effective Network Usage for Different Number of Jobs



Scalability tests show the economic models improving more than LFU as more jobs are added to the Grid.



Conclusion

- The **economic models** generally make more **efficient use of** Grid **resources** than traditional algorithms such as LFU.
- In particular situations the Economic models are considerably **faster** than LFU and **improve over the runtime** of the simulation.
- Future work:
 - Add simulation of CE clusters and hierarchical storage systems.
 - Extension of OptorSim **towards Grid Services**.
 - Simulate various Grid **failure modes** (network failures, unavailability of resources,...)
 - Deployment of algorithms into **Replica Optimization Service** and testing on the **real Grid**.

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