University of Glasgow

DEGREES OF M.Sc., M.Sci., M.Eng., B.Eng., B.Sc., M.A. and M.A. (Social Sciences)

COMPUTING SCIENCE M:
GRID COMPUTING

(Answer all of Section A and 2 out of 4 questions from Section B.)

Section A

1. **(a)** Explain what is meant by Grid Computing and discuss how a computational grid differs from a distributed computer system.

[4]

(b) Many computational grids are built using web services technology. Discuss why this technology is well suited to Grid Computing, and comment on why it might be succeeding in this role, when other remote procedure call technologies have failed to gain widespread acceptance.

[8]

(c) What are the limitations of web services technology for Grid Computing? Your answer should include an explanation of the specific limitations of web service protocols when applied to Grid Computing, and also a more general discussion of which features of computational grids are not addressed by the web services concept and must be supported by other protocols or technologies if they are to be used in a Grid.

[8]

(d) The programming assignment for this module used several Grid Computing technologies, including the Globus toolkit, the Condor scheduler, and the Permis security infrastructure. These were employed to retrieve biological sequence data from a remote database, and to build a computational grid to run large numbers of BLAST jobs in parallel. If you were given the job of rewriting or otherwise re-engineering these Grid Computing technologies to support a production-quality computational grid for running many parallel BLAST jobs on data retrieved from remote databases, what would you do differently? Discuss the changes you would make to the underlying technologies and libraries, and explain the reasons why you would make those changes.

[10]

Section B

2. **(a)** Explain the principles behind Public Key Infrastructures (PKIs) and outline their advantages and disadvantages with regard to grid security. Your answer should include a description of the different architectural choices that are possible when establishing PKIs and their associated trade-offs.

[9]

(b) Globus is one of the primary software toolkits for building grids today. Globus uses Globus Security Infrastructure (GSI) to support security. Explain the function of the grid mapfile in GSI.

[1]

(c) Discuss the advantages and disadvantages of the Globus GSI based approach for grid security?

[5]

(d) Shibboleth offers an alternative and more user-oriented approach to security. Explain why Shibboleth can be described as user oriented.

[1]

(e) Describe the advantages and disadvantages of Shibboleth for security in the context of the grid.

[4]

3. **(a)** Peer-to-peer systems can provide a distributed hash table (DHT) service, which may be used to lookup unstructured names and route queries to objects identified by those names. Explain how a DHT-based name service differs from the naming and resource discovery mechanisms used in Grid Computing systems based on web services. Discuss why a DHT-based resource discovery service might be useful in a Grid Computing system, and what benefit it might provide compared to a system implemented using web services.

[6]

(b) Describe how the Chord DHT algorithm works, and explain how it might be used to implement a distributed file system.

[6]

(c) Describe how you would make the distributed file system from your answer to part (b) robust to the failure of some nodes.

[2]

(d) You have read about the OceanStore distributed file system. This is built upon a DHT algorithm, and claims to provide a secure, fault tolerant and robust large-scale data storage infrastructure. To what extent to you think that systems like OceanStore are realistic? Discuss the implementation and deployment challenges, and the issues that should be considered before trusting important data to a system like OceanStore.

[6]

4. **(a)** There are many possible approaches to categorizing resource management problems and/or solutions in Grid computing. Identify one such perspective and give a brief taxonomy that addresses this point of view. You may base your answer on the survey by Krauter et al, or other readings, or your own practical experience with a range of Grid tools.

[4]

(b) Identify one particular resource and carefully explain why accounting for its use is complicated by the issue of autonomy.

[3]

(c) "Neither classical market pricing, nor congestion pricing, will be used in future Grid systems". Discuss.

[6]

(d) Assume that in a hypothetical future Grid toolkit, the issues of usability, accounting and contract agreement have all been resolved. Focusing on issues of fault tolerance and/or load balancing, identify the key technical problems that could influence a decision as to whether to use a multi-provider large-scale Grid, or to install a simpler but less powerful local system.

[7]

5. At a recent seminar, an eminent e-scientist asserted "There is no need for new protocols like FAST TCP to support bulk data transfer; an appropriately configured GRID FTP is sufficient to exploit IP networks with large bandwidth-delay products."

Do you agree? Support your answer by describing the main characteristics of end-to-end flow control in IP networks, and how FAST TCP and GRID FTP address these characteristics for bulk data transfer in IP networks with large bandwidth-delay products.

[20]