

Wednesday, May 9, 2007
2:30p.m. – 4:15p.m.

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 Grid Computing, and how it differs from traditional distributed systems. Discuss the extent to which these differences are due to politics and economics, rather than being caused by technical factors in the design of the systems.

[15]
- (b) You have used several Grid Computing technologies in the tutorials and programming assignment, including the Globus Toolkit and Condor scheduler. While these technologies are widely used, it has sometimes been suggested that they are heavyweight, over-complex, and fragile, and that the community should attempt to develop a simpler alternative. Do you consider this criticism fair? Is the complexity due to poor design and/or implementation, or is it inherent in the problem domain? Justify your answer.

[15]

Section B

2. (a) The Internet2 Shibboleth technology is currently being rolled out across UK academia. *Identity providers* and *service providers* underlie the Shibboleth approach and are used to support *federated access control* and *single sign-on* to distributed resources. Explain what is meant by each of the italicized terms. [4]
- (b) The UK e-Science community has adopted a single-tier *Public Key Infrastructure* (PKI) based upon a centralized *certification authority*, with regional *registration authorities* to support *single sign-on* to Grid resources. Explain the italicized terms and how they are used to securely access resources such as the UK e-Science National Grid Service. Outline the differences between single sign-on in the context of a PKI for the Grid, and single sign-on in the context of Shibboleth. [6]
- (c) Discuss the advantages and disadvantages of the Shibboleth approach to security over the PKI-based approach used by the UK e-Science community. Your answer should include a discussion of the dynamicity, scalability, usability and overall level of security of the two approaches. [10]

3. (a) Many Grid Computing systems are built using the web services framework to provide platform independent remote-procedure calls, on which grid services are layered. The underlying web services force a particular naming model, based on URLs to identify services, where each URL encodes information on the location of the service, and the access protocol. Discuss the advantages and disadvantages of embedding location and access information into service names. [4]
- (b) An alternative naming model might use unstructured labels to name grid services and resources, with a separate lookup service providing a mapping between the label and the location/access method for the resource. Describe how you might implement such a naming scheme in a scalable manner, explaining your design choices. [10]
- (c) We discussed the Map-Reduce algorithm in use at Google to perform queries across large data sets. This scales by removing the need to explicitly name data objects, instead providing an implicitly parallel solution for operating on large sets of objects. Discuss whether the Map-Reduce concept can be adapted for use in a heterogeneous multi-organizational Grid Computing environment (and how such adaptation would affect its implementation and performance), or whether it is limited to operation in homogeneous cluster computing environments. [6]

4. (a) The National Health Service has seen an alarming increase in asthma attacks in various locations within the United Kingdom, as reported by doctors' surgeries. The NHS has decided to adopt a high-tech approach to monitoring asthma attacks and correlating those attacks with location and other data. The key component of this system is an inhaler fitted with:
1. A battery
 2. Circuitry to note the time of each use of the inhaler
 3. Sufficient flash memory to store this time stamp data for the maximum number of doses contained in the inhaler
 4. A radio to enable communication of the time stamp data to other systems (this radio has a range of 10 meters)

Each doctors' surgery, NHS hospital, and ambulance is equipped with one or more systems that possess the same type of radio and a GPS receiver; each such system constantly looks for passing inhalers. If detected, a connection is established between the inhaler and the GPS-equipped system, and the inhaler transmits its accumulated time stamp data to the GPS-equipped system; if the transmission is successful, the inhaler zeroes its buffer of accumulated time stamps. The GPS-equipped system, upon successful reception of the time stamp data, associates this data with its current GPS location, and then forwards this data onto a central collection system at an NHS site in Manchester. This geographical and timing information can then be analysed along with other data associated with particular locations, such as pollen count, temperature, and humidity.

You have been tasked with constructing the networking protocols required to achieve this functionality. Describe how you would solve this problem using delay-tolerant networking protocols. Be sure to address the following points in your answer:

1. Why is this problem amenable to a solution using delay-tolerant networking?
2. What additional information is required in the inhaler to make this work?
3. How many different Internets are required and why?
4. Health data is subject to severe privacy and need-to-know constraints; what is required to maintain the privacy of the data being transmitted?

Do not be afraid to draw pictures/diagrams if it will aid your discussion.

[20]

5. (a) Condor uses a centralized scheduler; Globus provides a decentralized scheduling infrastructure. Explain why the overall system architectures and related design decisions have led to the chosen scheduler architecture in each of these two systems, and in any one other system of which you are aware. [6]
- (b) Why is periodic push widely used for dissemination of dynamic resource information in Grid systems? Suggest, and justify, a scenario in which an alternative approach might be appropriate. [5]
- (c) Briefly describe (i) shadow pricing, and (ii) Vickrey auctions. [4]
- (d) Either:
- Briefly describe two Grid-focused workflow management systems. Identify one key characteristic that distinguishes the two, explaining its significance as a feature of the systems.
- Or:
- Explain the significance of negotiation in Grid resource allocation and briefly describe two issues that must be addressed when designing a negotiating mechanism. [5]