

**This question paper consists  
of 4 printed pages, each  
of which is identified by the  
Code Number COMP3211.**

**This is an open book examination.  
Any written or printed material is permitted.**

**© UNIVERSITY OF LEEDS**

School of Computing

**January 2018**

**COMP3211**

Distributed Systems

Answer all THREE questions

Time allowed: 2 hours

**Question 1**

- (a) When distributed systems are designed and engineered, certain fundamental properties have to be taken into account, including:

- Concurrent execution of components
- Independent failure modes
- Communication delay
- No global time.

Discuss briefly the implications of these properties (separately or in combination) on the engineering of large-scale, widely distributed systems. **[4 marks]**

- (b) Consider the Web search engine introduced in the lectures in the context of a three-tier application. A single data request message is sent by the client to the single application server. The application server uses the content of this message to create a pair of request messages to send to two databases. Responses to these messages are returned from the two databases to the application server and are forwarded on by the application server as two separate responses to the client.

- (i) Explain the transparency implications of this design decision. **[3 marks]**

- (ii) What would be the consequences of combining the two responses (one from each database) into a single reply to the client? **[3 marks]**

- (c) A non-real time compute-intensive task is run three times in a system, at different times of the day. The run times are 60, 61, and 80 seconds, respectively. Given that the task performs exactly the same computation each time it runs, how do you account for the different run times? **[3 marks]**

- (d) A student is overheard saying the following regarding distributed systems:

- (i) To generate a WSDL document, run the `wsimport` tool followed by the name of the `.java` file containing the class. **[1 mark]**
- (ii) UDDI is a true discovery service. **[2 marks]**
- (iii) UDDI and LDAP are just different implementations of the same concept. **[2 marks]**
- (iv) NTP has a linear design and is based on the principles of having all machines get as close as possible to UTC. **[2 marks]**

Comment briefly on these statements, correcting any inaccuracies or omissions in them.

**[question 1 total: 20 marks]**

**Question 2**

- (a) Consider a message queue system. In order to automatically start a process to fetch messages from an input queue, a daemon is often used that monitors the input queue. Propose a simple, alternative implementation that does not make use of a daemon. **[3 marks]**
- (b) In 221 B.C., Emperor Qin unified several formerly warring states into a new country, which we now call China. Perhaps one reason China has endured as a nation was Qin's introduction of standards, which consolidated cultures and facilitated trade: a standard distance for wheels on carts which allowed them to travel efficiently on any road, a common written language that everybody could use to exchange messages (even if they did not speak the same language), and strong defenses against outside attacks (like the Great Wall of China). How does this statement relate to SOAs and Web Services? **[4 marks]**
- (c) You are asked to build a Web services client application that integrates an Amazon SOAP-based Web service and a Google RESTful search Web service. What are the factors that may affect the performance of your application? **[3 marks]**
- (d) A client and a server need to synchronise their clocks in a distributed system.
- (i) The client's clock reads 5:26:00. The server's clock reads 5:14:00 when they synchronize using the Berkeley algorithm. Assume message delays are negligible. What is the time at the client after synchronisation? Note: the time format is HH:MM:SS. **[2 marks]**
  - (ii) The client's clock reads 5:26:08. The server's clock reads 5:16:44 when they synchronize using Cristian's algorithm. Assume RTT is 2 seconds. What is the time at the client after synchronization? Note: the time format is HH:MM:SS. **[2 marks]**
- (e) Many distributed algorithms require one process to act as coordinator, at least initially.
- (i) If a process detects that the original coordinator is no longer responding to requests, it initiates a new election. Suppose that two processes detect the demise of the coordinator simultaneously and both decide to hold an election using the bully algorithm. Would an election be accomplished successfully? Explain your answer. **[3 marks]**
  - (ii) Using the bully algorithm, describe the election amongst a group of 6 nodes labelled 0..5. Initially node 5 is the coordinator but it crashes. Node 2 is the first node to notice that the coordinator has crashed. Your answer should define the types of messages exchanged between nodes. It should also clearly indicate all messages exchanged between nodes in the election process and which node is elected as the new coordinator. **[3 marks]**

**[question 2 total: 20 marks]**

**Question 3**

- (a) From a local machine `mycomputer.comp.uleeds.uk`, you would like to access the URL `www.maps.google.com/home/index.html`. Assume that your local name server is the one covering the `comp` domain, and that the `uleeds` and `uk` are two other administrative zones. There is a root server `R` available, and `com`, `google`, and `maps` are administrative zones for the the URL you are trying to access. Draw a diagram showing the recursive name resolution process, assuming that no part of the name is cached at any of the servers. Number your steps, and at each step, show the information that is being passed between the name servers. **[4 marks]**
- (b) Which statement best describes the relationship between application, server, and client in a multi-tenancy cloud environment? **[2 marks]**
- (i) Single instance of software running on a server and serving one client.
  - (ii) Single instance of software running on a server and serving multiple clients.
  - (iii) Multiple instances of software running on a server and serving multiple clients.
  - (iv) Multiple instances of software running on multiple servers and serving one client.
- (c) MapReduce is a programming model and an associated implementation for processing and generating large data sets. For each of the following say whether the statement is true or false and justify briefly your answer.
- (i) Each mapper must generate the same number of key/value pairs as its input had. **[2 marks]**
  - (ii) Reducer is applied to all values associated with the same key. **[2 marks]**
  - (iii) Is it possible to start reducers while some mappers still run. **[2 marks]**
- (d) What are the main characteristics of a distributed file system in a large-scale cloud system? Illustrate your answer using Google File System (GFS) or HDFS. **[3 marks]**
- (e) As an expert in distributed systems, you have been contacted by Leeds City Council to lead a project on monitoring environmental parameters in the city, which is experiencing high pollution coming directly from road traffic. The main goal of the project is to achieve sustainable management of the traffic in the city by using two key-elements: a pervasive air-quality sensors network connected to the cloud as well as prediction models. Propose a cloud computing-based solution that considers the following:  
1) Pervasive air-quality sensors network; 2) Edge computing model with connected devices; 3) Data collection; 4) Parameters to measure; 5) ZigBee/Wi-fi communication, and 6) Prediction models.
- Discuss how to deploy your solution and include any information that you feel is relevant to back it up. **[5 marks]**

**[question 3 total: 20 marks]**  
**[grand total: 60 marks]**