CARDIFF UNIVERSITY
EXAMINATION PAPER

Academic Year: 2017-2018

Examination Period: Spring

Examination Paper Number: CMT202

Examination Paper Title: Distributed and Cloud Computing

Duration: TWO hours

Do not turn this page over until instructed to do so by the Senior Invigilator.

Structure of Examination Paper:

There are **THREE** pages.
There are **FOUR** questions in total.
There are no appendices.
The maximum mark for the examination paper is **60 marks** and the mark obtainable for a question or part of a question is shown in brackets alongside the question.

Students to be provided with:

The following item of stationery is to be provided:
ONE answer book.

Instructions to Students:

Answer **THREE** questions.

Important note: if you answer more than the number of questions instructed, then answers will be marked in the order they appear only until the above instruction is met. Extra answers will be ignored. Clearly cancel any answers not intended for marking. Write clearly on the front of the answer book the numbers of the answers to be marked.

The use of a translation dictionary between English or Welsh and another language, provided that it bears an appropriate school stamp, is permitted in this examination.

1 (a) Define what is meant by the term *distribution transparency*. As a means of communication, a WebSocket does not offer *location transparency*. Justify this statement. **[5]**

(b) Consider the problem of counting the number of occurrences of each word in a large collection of documents. Provide pseudocode for *map* and *reduce* functions of a MapReduce solution to this problem. **[5]**

(c) Explain the difference between a *software architecture* and a *system architecture* in the context of a distributed system. Provide an example of a software architecture style and an example of a system architecture style commonly used in distributed systems. **[5]**

(d) Explain the difference between a *process* and a *thread*. Provide one advantage of using threads, as opposed to processes, in the context of a distributed system. **[5]**

2 (a) Describe what is meant by the term *code migration*. Provide a real-world example of how code migration can improve performance in a distributed system. **[5]**

(b) Describe how *Message Oriented Middleware* (MOM) achieves persistent and asynchronous communication. Give an example of how performance in a distributed system may be improved using asynchronous communication. **[5]**

(c) Describe how an *asymmetric (public-key) cryptosystem* functions. How can such a cryptosystem be used to perform message authentication? **[5]**

(d) Describe the purpose of *access control*. Describe how access control can be achieved using an access control matrix. **[5]**

3 (a) *Peer-to-peer* (P2P) systems fall into three main categories: *centralized systems, brokered systems* and *decentralized systems*. Explain how these three categories of system differ. **[6]**

(b) In *unstructured peer-to-peer systems*, significant improvements on search results can be provided by the adoption of particular search strategies. Name and briefly describe two alternative search strategies to *simple flooding*. **[4]**

(c) *Napster* is an example of a brokered system, while *Gnutella* (version 0.4) is an example of a decentralized system. Explain, for each of the two systems:

(i) How a newly joined node publishes the files it wishes to make available.

(ii) How a node finds the location of a desired file.

[6]

(d) Another example of distributed technology is *blockchain*. Briefly explain the terms *blockchain* and *smart contract*. **[4]**

4 (a) What are the five main principles of Cloud computing? **[5]**

(b) Let us assume we have services that we would like to deliver to some specific geographical locations. Which virtualization technology (VM or container technology) would you use to make them available rapidly and on-demand? Justify your choice. **[4]**

(c) Cloud computing services such as *Amazon Elastic Compute Cloud* (EC2) assign virtual machines (VMs) to users instead of allocating physical machines directly. Doing so provides at least four major benefits to Amazon. Explain what these four benefits are, giving a brief motivation for each one. **[8]**

(d) List *three* differences between Virtual Machines and Docker containers. **[3]**