One and a half hours

UNIVERSITY OF MANCHESTER
SCHOOL OF COMPUTER SCIENCE

Fundamentals of Distributed Systems

Date: Thursday 21st May 2015
Time: 09:45 - 11:15

Please answer any TWO Questions from the THREE Questions provided.

Use a SEPARATE answerbook for each QUESTION.

This is a CLOSED book examination

The use of electronic calculators is permitted provided they are not programmable and do not store text
1. a) In this course unit, we have divided the reasons for distributing systems into two types.

   i) Consider comparison websites (e.g., for hotel bookings, for insurance). State which of the two types they are an example of and give a brief explanation of your reasons for thinking so. (3 marks)

   ii) The use of cloud computing in its software-as-a-service view (e.g., Google Docs) exemplifies both types of reasons for distributing systems. Briefly explain why this can be claimed. (4 marks)

b) This course unit has used the following definition of a distributed system:

   A distributed system is a system in which independent, self-sufficient, often heterogeneous and autonomous, spatially-separated components must use a common interconnect in order to exchange information, with a view to coordinating their actions and, thereby, cooperate to achieve coherent, consistent goals and purposes in the eyes its users.

   i) In a distributed system, “failures occur and we may not know or be told”. Briefly explain how this follows from the definition above and the axioms/fallacies of distributed computing discussed in the course. (3 marks)

   ii) In a distributed system, “asynchrony (i.e., non-aligned timelines) is unavoidable”. Briefly explain how this follows from the definition above and the axioms/fallacies of distributed computing discussed in the course. (4 marks)

c) i) The domain name system (DNS) provides a mapping from symbolic names to IP addresses based upon a globally distributed and hierarchical database. State which transparency goal is made significantly more achievable by the DNS and briefly explain the reason(s) behind this. (4 marks)

ii) The use of interoperable standards makes which transparency goal significantly more achievable? (1 mark)

d) i) Briefly explain what the functionality of a memory-resident batch monitor in the mainframe era was, and what problem (in terms of relative costs) it was a solution for. (4 marks)

ii) Briefly explain why the first personal computers came with rudimentary OSs compared to those available for mainframes. (2 marks)
2. a) i) Briefly explain what concept in distributed systems is exemplified by a notice board (i.e., a place where people leave messages for other people to see and perhaps respond to). (2 marks)

ii) Suppose that whenever you try to leave a message in a notice board you always find it has no space left. Briefly explain what problem is exemplified by this situation. (3 marks)

b) Imagine you are a software engineer in a company that is planning a move towards using web-based distributed computing. The technical manager has called for a discussion of the challenges involved. The fact that the existing systems were originally designed with a view to being CPU-efficient causes her particular concern. She asks you to comment on whether her concerns are well-founded. State whether you agree with your technical manager and briefly explain why. (4 marks)

c) Briefly explain (in terms of the architectural paradigms for distributed computing covered in this course unit) why peer-to-peer architectures (P2P) are said to lead naturally to balanced loads and graceful scaling. (4 marks)

d) i) In interprocess communication, there are four basic forms of ‘send’. They are: blocking and asynchronous, blocking and synchronous, non-blocking and asynchronous, non-blocking and synchronous. State which forms of ‘send’ are not possible in the ‘receive’ case and briefly explain why. (1 mark)

ii) Imagine a notice board (i.e., a place where people leave messages for other people to see and perhaps respond to). Among the basic forms studied in the course unit, state which are the forms of ‘send’ and ‘receive’ that are used in this notice board scenario, making sure that you explain the main points of the analogy. (3 marks)

e) In a master-workers approach to parallelization/distribution of load, state which are the three primitive operations that the master must carry out and briefly describe them. (3 marks)
f) When the course unit introduced the map-reduce approach, an example was given of how a mapper and a reducer function could be written that would count the occurrences of words in Web documents. The mapper and reducer function below (in pseudocode) are structurally very similar yet they compute something different. Briefly explain what they compute. (You do not need to describe how the computation is carried out in the underlying map-reduce engine.)

mapper(String input_key, String input_value):
    # input_key: document name
    # input_value: document contents
    for word w in input_value:
        EmitIntermediate(w[0:1], length(w))

reducer(String output_key, Iterator intermediate_values):
    result = 0
    for v in intermediate_values:
        if v > result:
            result = v
    Emit(result)

(4 marks)

g) Briefly explain why it is claimed that, in the WS-* view of services on the Web, HTTP is just a transport mechanism.

(1 mark)
3. a) State whether you agree or disagree with the statement that ‘a cache is a remotely-residing data storage area, one of whose effects is to directly improve latency’. If you agree, explain why. If you disagree, restate the claim so it is correct and explain the changes you have made in the process of correcting it. (4 marks)

b) Imagine you are a software engineer in a web-based company. You are part of a team that aims to build a browser-based system that will not be dependent on users allowing cookies to be set by the browser they are using. A colleague of yours said: “This goal is impossible to achieve. Everybody knows that HTTP is a stateless protocol.”. The technical manager turns to you for a second technical opinion.

   i) State whether you agree with your colleague about the goal being impossible. (1 mark)

   ii) Briefly explain what relationship between the goal and HTTP statelessness your colleague is relying on. (2 marks)

   iii) Briefly explain your reasons for agreeing or disagreeing with your colleague. (3 marks)

c) i) State the purpose of the two-phase commit protocol (2PC). (2 marks)

   ii) Describe the steps in 2PC. (6 marks)

   iii) State which pre-requisite of 2PC can be satisfied by another protocol/algorithmdescribed in the course and state which protocol/algorithmin that is. (2 marks)

d) Draw a diagram representing the state of a circular distributed hash table after the insertion of the list of items

\[ L = [12, 130, 32, 4, 6, 5, 122] \]

using as hash function

\[ h(l) = l \text{ modulo } 8 \]

Your diagram should represent the keyspace by showing the keys as labels of the nodes represented as small circles and you should indicate for each node which items it is responsible for at the end of the insertion process. (You do not need to show the detail of the step-by-step progress through the list of items, just the final state.) (5 marks)