UNIVERSITY OF EDINBURGH

FACULTY OF SCIENCE AND ENGINEERING

COMPUTER SCIENCE FOUR

DISTRIBUTED SYSTEMS

Tuesday 18 April 2000

2:00 to 3:30

Examiners:

Prof. K. Turner (External)

Prof. G. Brebner (Chairman)

INSTRUCTIONS TO CANDIDATES

Answer any TWO questions.

All questions carry equal weight.

1. (a) The Chandry-Misra-Haas (CMH) distributed algorithm for deadlock detection uses a *waits-for* graph and *probe* messages. A probe is a triple (i, j, k) denoting that process P_i initiated the deadlock detection and the probe was sent by process P_j's site to process P_k's site.

   i. With the aid of a small example, explain how the CMH algorithm detects and breaks deadlocks. [*8 marks*]

   ii. How might the CMH algorithm be adapted to avoid detecting 'phantom' deadlocks and to avoid re-discovering deadlocks? [*7 marks*]

   (b) Consider the following technique for detecting distributed deadlock:

   > Freeze activity at regular time intervals. Once sites have frozen activity, they pass all information about locking and blocking to a deadlock detection manager which applies a centralised deadlock detection technique to the information.

   Under what circumstances would the CMH algorithm provide performance that is superior to this technique? Your answer should take into account characteristics of the workload as well as characteristics of the network. [*10 marks*]

2. (a) Fault tolerance is the ability of a system to continue to behave in a well-defined manner even though faults have occurred. Two limited forms of fault tolerance are:

    non-masking fault tolerance (which ensures liveness, but not safety) and

    fail safe fault tolerance (which ensures safety, but not liveness).

    Give two examples of distributed systems applications in one of which non-masking fault tolerance is preferable to fail safe fault tolerance, and in the other, fail safe is preferable to non-masking. [*5 marks*]

   (b) The *Network Time Protocol* (NTP) is an Internet standard for synchronizing computer clocks.

    i. Describe NTP's technique for measuring the delay and offset between two computers that communicate via message passing. [*5 marks*]

    ii. Using a worked example, explain why this technique may not be reliable if, for a particular pair of communicating computers, the message propagation delay tends to be longer in one direction than the other. [*5 marks*]

   (c) In a remote procedure call (RPC), a process executing on a client requests the execution of a procedure on a remote server.

    i. Consider a distributed system in which there is an RPC service and there exists the possibility of server and network failures. What are the difficulties of providing **exactly-once** semantics to the RPC service, i.e. ensuring that each client RPC invocation is executed exactly once on the appropriate server? [*5 marks*]

    ii. Briefly describe the advantages of XMLRPC, as compared to distributed object systems such as CORBA. [*5 marks*]

3. (a) For protocols that require a leader, it is sometimes desirable to provide the facility to hold a leader election in case the current leader fails.

   i. In the 'Bully' leader election algorithm, each site, i, has safe storage for a structure, $S(i)$, which contains three elements, namely
   $S(i).d$: the task that site i is performing.
   $S(i).c$: the site number of the coordinator.
   $S(i).s$: the state of site i.
   With reference to this structure, S, describe the Bully algorithm. [*10 marks*]

   ii. What are the main assumptions that the Bully algorithm makes about the reliability of the distributed computing environment? [*5 marks*]

   iii. What goes wrong with the Bully algorithm if these assumptions do not hold? [*5 marks*]

   (b) With reference to an existing commercial distributed system, give two examples of techniques that might be used to improve the performance and robustness of a leadership election protocol. [*5 marks*]