

NANYANG TECHNOLOGICAL UNIVERSITY**SEMESTER 2 EXAMINATION 2022-2023****CE4013/CZ4013/SC4051 – DISTRIBUTED SYSTEMS**

Apr/May 2023

Time Allowed: 2 hours

INSTRUCTIONS

1. This paper contains 8 questions and comprises 6 pages.
 2. Answer **ALL** questions.
 3. This is a closed-book examination.
 4. Questions do not carry equal marks.
-

1. (a) An information server maintains the COVID-19 situation reports of different countries and updates them daily. The server would like to keep interested clients informed of the latest report through callback. Design a Java remote interface for the clients to allow the server to perform callback, and a Java remote interface for the server to allow the clients to register and deregister their interests in the report of a specified country. Assume that the reports and country names are represented by strings.

(5 marks)

- (b) A client makes remote method invocations to a server. The client takes 3 milliseconds to compute the arguments for each request, and the server takes 20 milliseconds to process each request. The local operating system processing time for each send or receive operation is 1 millisecond, and the network propagation time to transmit each request or reply message is 4 milliseconds. Marshalling or unmarshalling takes 1 millisecond per message. The server runs two threads on two processors, and the client runs two threads on a single processor. Assume that context-switching times are negligible. Calculate the time taken by the client to generate and return from two remote method invocation requests.

(6 marks)

2. A client C in NTU would like to resolve a name “www.ox.ac.uk”. List the sequence of all messages sent among C and the relevant name servers in the name resolution process, assuming name mappings are not cached at non-authoritative name servers initially. In your answer, use the format “NS(ddd)” to indicate the authoritative name server for the domain “ddd”, and use the format “ $C \rightarrow \text{NS}(ddd)$ ” to indicate a message sent by C to NS(ddd). (6 marks)

3. A distributed file system consists of a server and two clients A and B . Let f be a single-block file stored at the server. Figure Q3 shows the operations performed by clients A and B on file f , where “ o ”, “ r ”, “ u ” and “ c ” represent open, read, update and close operations respectively. A and B both start with an empty client cache, and the caches are large enough to avoid any replacement. The updates made by B do not change the length of file f . The transmission delay in the network and the processing times at the server and clients are negligible.

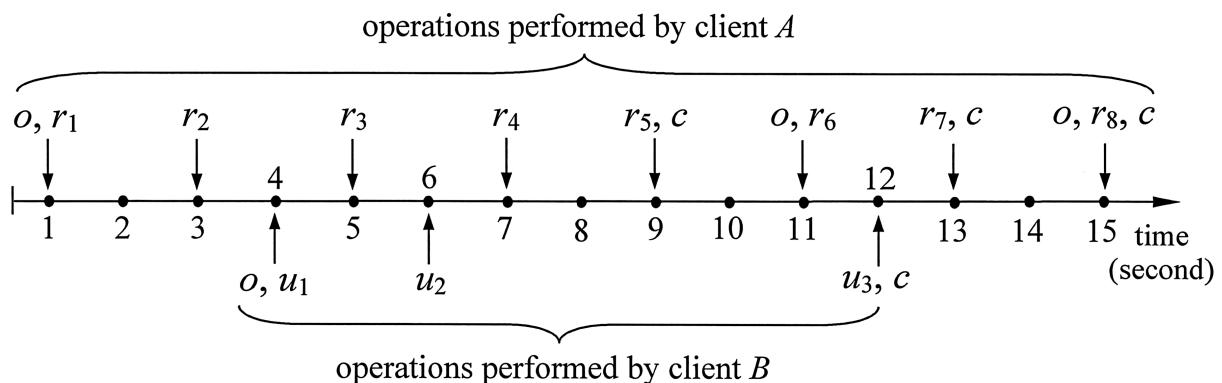


Figure Q3

For each case below, determine the time instants when A needs to contact the server, and determine which read operations of A return the up-to-date file.

- (a) The system is an NFS. The clients use a freshness interval of 3 seconds to maintain cache consistency, and client B uses a bio-daemon process to send its updates to the server right after making each update. (6 marks)
- (b) The system is an AFS that implements session update semantics. (6 marks)

4. In a peer-to-peer file sharing system based on Chord routing, the identifier circle ranges from 0 to 127 (i.e., $2^7 - 1$). There are 9 nodes in the system: N1, N8, N21, N30, N36, N51, N91, N97 and N116, where the numbers behind ‘N’ denote the node identifiers on the identifier circle.

- (a) What is the route of a query issued by node N51 for the location information of file K34 (the number behind ‘K’ denotes the key identifier of the file on the identifier circle)?

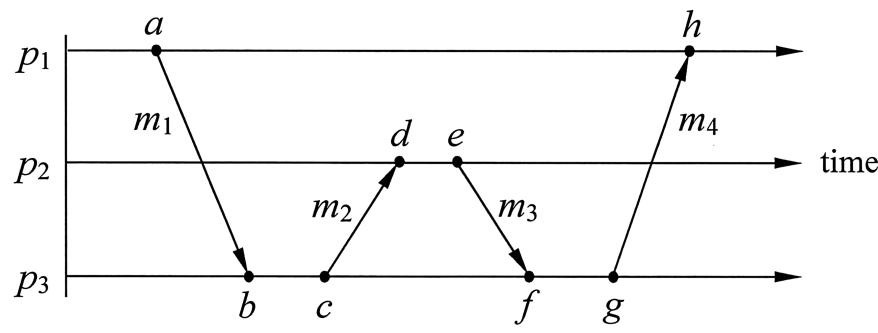
(5 marks)

- (b) When a new node joins the system, nodes N8 and N30 both need to update their finger tables. What are the possible node identifiers of this new node? (Note: besides N8 and N30, there may be other nodes which also need to update their finger tables due to the new node.)

(6 marks)

5.

- Three processes p_1 , p_2 and p_3 are running on three different computers in an asynchronous distributed system. Figure Q5 shows the messages m_1 to m_4 sent between the processes, where events a to h represent the sending and receiving events of these messages.

**Figure Q5**

- (a) Suppose that the transmission delay of each message is at least 50 ms. Clock drifts are negligible. Let t_x be the local clock reading at event x in the process where x occurs (for example, t_a is the local clock reading of p_1 at event a , and t_b is the local clock reading of p_3 at event b). Assume that p_1 knows the clock readings t_e and t_g , and also knows that $t_h - t_a = 1000$ ms, $t_c - t_b = 100$ ms, $t_f - t_c = 500$ ms, $t_g - t_f = 100$ ms, $t_e - t_d = 300$ ms.

- (i) If p_1 would like to synchronize its local clock with p_3 's clock as accurately as possible, what time should p_1 set its clock to when it receives message m_4 from p_3 ? What is the accuracy of this setting?

(5 marks)

Note: Question No. 5 continues on Page 4

- ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library
- (ii) If p_1 would like to synchronize its local clock with p_2 's clock as accurately as possible, what time should p_1 set its clock to when it receives message m_4 from p_3 ? What is the accuracy of this setting?
(5 marks)
- (b) Timestamp all the events in Figure Q5 using the vector clocks.
(5 marks)
- (c) Assume that message delivery on each unidirectional point-to-point channel follows FIFO order. Process p_3 initiates the Chandy-and-Lamport algorithm immediately after event c to record a snapshot of the system. List all the possible snapshots finally recorded. Indicate clearly the process states and channel states in the snapshots. In your answer, use S_1 , S_2 and S_3 to represent the initial states of processes p_1 , p_2 and p_3 respectively, and use S_x to represent the state of the process (where event x occurs) immediately after event x 's occurrence (for example, S_a is the state of process p_1 immediately after event a occurs, and S_b is the state of process p_3 immediately after event b occurs).
(5 marks)
6. A replicated shared object service hosts an integer object x whose initial value is 0. Table Q6 shows the operations performed by two clients on x . For example, $\text{read}(x) \rightarrow a$ denotes a read operation on x returning a value a , and $\text{write}(x, 2)$ denotes a write operation setting the value of x to 2. The operations performed by each client are listed in the order that they are performed by the client.
- Table Q6**
- | Program order | Client A | Client B |
|---|--|--|
| <div style="position: absolute; left: 50%; top: 50%; transform: translate(-50%, -50%);"> \downarrow </div> | $\text{write}(x, 1)$
$\text{read}(x) \rightarrow a$
$\text{write}(x, 2)$
$\text{read}(x) \rightarrow b$ | $\text{write}(x, 3)$
$\text{read}(x) \rightarrow 1$
$\text{write}(x, 4)$
$\text{read}(x) \rightarrow c$ |
- $1, 2, 4$
 $1, 2, 2$
 $1, 4, 4,$
 $4, 2, 1$
 $4, 2, 4$
- (a) If the object service is sequentially consistent, what are the possible combinations of the values a , b and c ?
(5 marks)
- (b) If the object service is FIFO consistent, what are the possible combinations of the values a and b ?
(5 marks)

Note: Question No. 6 continues on Page 5

- 3
7
D
1
(2 marks)
- (c) Suppose that there are four replicas of object x .
- (i) What are the maximum numbers of crash failures and arbitrary failures that can be tolerated respectively under passive replication?
7
7
(2 marks)
- (ii) What are the maximum numbers of crash failures and arbitrary failures that can be tolerated respectively under active replication?
7
7
(2 marks)
- (iii) Let R be the minimum allowable number of replicas in a read quorum and W be the minimum allowable number of replicas in a write quorum. List all possible combinations of R and W permitted by Gifford's quorum consensus scheme to handle network partitions.
(3 marks)
7. Three processes p_1, p_2 and p_3 are running in an asynchronous distributed system. The vector clock mechanism is used to timestamp the events that occur in these processes, including the events of sending messages, receiving messages and the internal actions carried out by the processes. Suppose that two events a and b have timestamps (10, 20, 30) and (40, 50, 60) respectively.
- (a) What is the maximum possible number of events x in the system satisfying both $a \rightarrow x$ and $x \rightarrow b$ in causal ordering? Briefly explain your answer and draw a diagram to illustrate such a scenario.
(5 marks)
- (b) What is the minimum possible number of events x in the system satisfying both $a \rightarrow x$ and $x \rightarrow b$ in causal ordering? Briefly explain your answer and draw a diagram to illustrate such a scenario.
(5 marks)
8. (a) In an asynchronous distributed system, two processes A and B are connected by a bidirectional communication channel. Each process wants to select and take one of two actions x and y . Assume that the processes suffer from crash failures, but the communication channel is reliable and does not fail. Can a protocol be devised for the two processes to agree upon the action to take? Briefly explain your answer.
(4 marks)

Note: Question No. 8 continues on Page 6

- (b) In the Ricart-and-Agrawala algorithm for mutual exclusion, instead of using logical clock readings to timestamp requests, suppose that every process maintains a monotonically increasing counter to record how many requests it has made, and uses the counter readings to timestamp its requests. With this modification, does the algorithm still guarantee that at most one process may execute in the critical section at any time? Briefly explain your answer.

(9 marks)

CE4013 DISTRIBUTED SYSTEMS

CZ4013 DISTRIBUTED SYSTEMS

SC4051 DISTRIBUTED SYSTEMS

Please read the following instructions carefully:

- 1. Please do not turn over the question paper until you are told to do so. Disciplinary action may be taken against you if you do so.**
2. You are not allowed to leave the examination hall unless accompanied by an invigilator. You may raise your hand if you need to communicate with the invigilator.
3. Please write your Matriculation Number on the front of the answer book.
4. Please indicate clearly in the answer book (at the appropriate place) if you are continuing the answer to a question elsewhere in the book.