2022 EXAMINATIONS

Lancaster University

Part II

COMPUTING AND COMMUNICATIONS – On-line Assessment [150 Minutes]

SCC.311 Distributed Systems

---

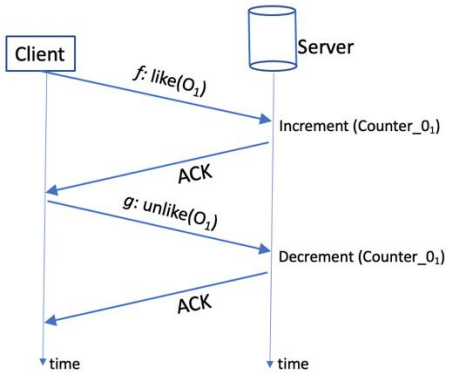
*Candidates are asked to answer **THREE** questions from **FOUR**; each question is worth a total of 25 marks.*

[Please turn over]

Question 1

1. a) Consider a like-tracking system for a social media application, where clients send `like' and `unlike' requests for content objects (e.g., images or videos), while the tracking server maintains a counter for each object to keep track of the number of likes each object received.

In a simple client-server implementation of this system, users (as clients) make *like(Object o)* and *unlike(Object o)* remote procedure calls. A *like(O_1)* call triggers an increment on the counter state for O_1, while an *unlike(O_1)* call triggers a decrement. The server responds to each incoming request with an Acknowledgement (i.e., ACK) after updating the counter as shown in the Figure. The server does not maintain any other state for a content object other than a counter.



(i) Explain why the like and unlike operations (as described above) are not idempotent. *(1-2 sentences)*

[2 marks]

(ii) Propose the necessary modifications to the system to achieve idempotent like/unlike operations. *(At most 5-6 sentences)*

[4 marks]

iii) A read(O_1) remote procedure call is introduced to the like-tracking system, which returns the current like count for the object O_1. Explain how your system with idempotent like/unlike requests computes the like count for a given object. *(1-2 sentences)*

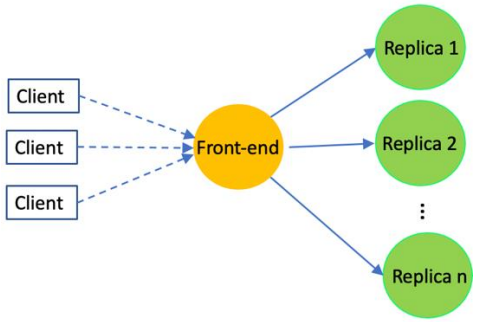
[2 marks]

[Please turn over]

1.b) Describe the necessary adjustments to the system to ensure that a user can only remove (unlike) its own likes from the objects and not the likes of other users. Make sure your system is not vulnerable to eavesdropping and replaying attacks. You can assume that the server already knows either the symmetric or the public key of each user. *(At most 4-5 sentences)*

[4 marks]

1.c) In a like-tracking system with replication, the like counters for the objects are replicated across *n servers*, and the clients send their RPC requests directly to a front-end server as shown in the Figure below. Describe a lightweight mechanism for the system to achieve total ordering of client requests at the replicas, assuming a reliable network between the front-end and the replicas. *(2-3 sentences)*



[2 marks]

1.d) Consider again the replicated like-tracking system above, where the front-end broadcasts read(), like(), and unlike() requests to all the n replicas; however, this time, the requests sent from the front-end to the replicas can be randomly dropped due to network errors. Imagine a scenario where a user Bob sends a like(O_1) request to the front-end which is broadcasted and received by w (out of n) replicas. Subsequently, a read(O_1) message from a user Alice is received by r replicas who each successfully send a response message with a like count.

(i) What is the minimum value of "r + w" for the front-end to receive at least one up-to-date like count for the request of Alice? (1 sentence)

[3 marks]

(ii) Justify your choice of the minimum r+w value. (1-2 sentences)

[3 marks]

[Please turn over]

1.e) In the system described in e), some updates (such as like() and unlike()) may be missing from some of the replicas at any given time due to network errors. Describe the necessary extensions to the system for the front-end to be able to identify the up-to-date response(s) to a read() request that are returned by the replicas. (Up to 4-5 sentences)

[5 marks]

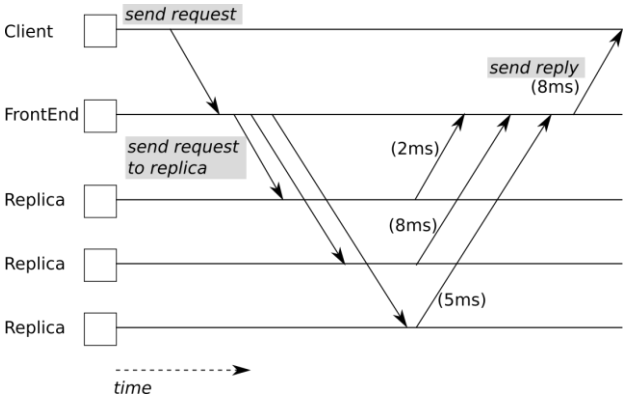
Total 25 marks

[Please turn over]

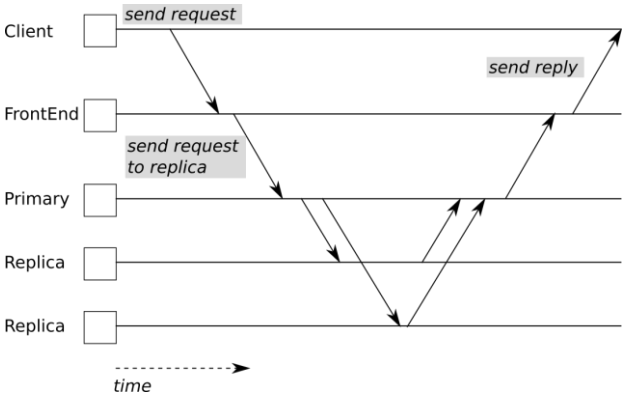
Question 2

2.a Consider the below two diagrams on different replication styles (i) and (ii).

(i)



(ii)



(i) Name replication style (i):

[2 marks]

(ii) Name replication style (ii):

[2 marks]

Consider the following list of possible services: [Personal File Storage], [Flight Booking], [Internet Banking], [An Online Discussion Forum].

(iii) Name **one** of the above services that would be suitable for replication style (i), and briefly justify your answer.

[3 marks]

(iv) Name **one** of the above services that would be suitable for replication style (ii), and briefly justify your answer.

[3 marks]

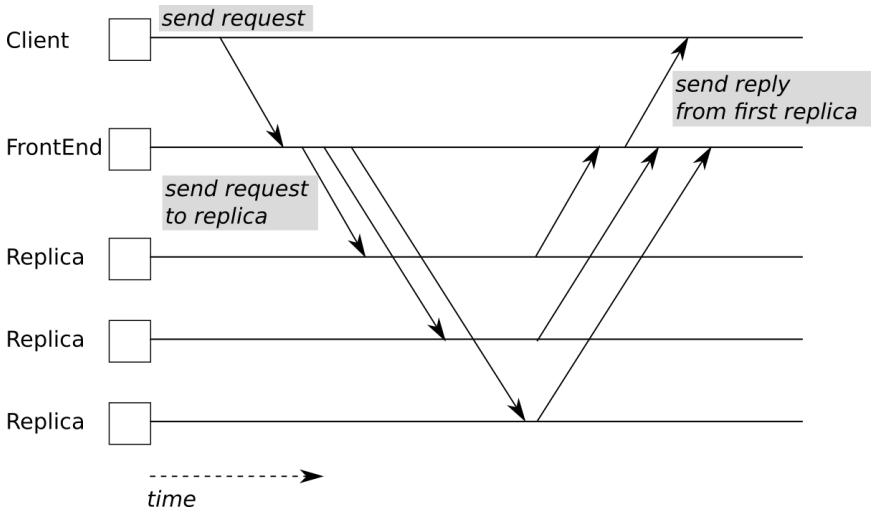
[Please turn over]

2.b Replication style (i), above, is annotated with round-trip delay times for messages. Assuming that compute time is negligible at all points, how long would a client wait to receive a reply in this case? Briefly explain your answer.

[4 marks]

2.c Illustrated below is an alternative implementation of replication style (i), which is intended to provide faster replies to client requests in order to improve quality-of-experience. Is this implementation correct? Justify your answer.

[4 marks]



2.d In general, which replication style, from (i) or (ii), would tend to be faster at processing client requests? Explain your answer.

[4 marks]

2.e. In replication scheme (i), if our system assumptions are that we should be able to tolerate up to 4 replicas exhibiting Byzantine faults, including network transport corruption, how many replicas in total would be needed to achieve consensus on each operation? Explain your answer.

[3 marks]

Total 25 marks

[Please turn over]

Question 3

3) a) Bob proposes the following total order broadcast mechanism where one node among N nodes is designated as the leader. Each node sends its messages first to the leader which then broadcasts the messages to all the nodes using First-In-First-Out (FIFO) broadcast.

(i) Is Bob's broadcast mechanism sufficient to achieve a total ordering of messages across all the N nodes ? (1 word)

[2 marks]

(ii) Explain the drawbacks of Bob's broadcast mechanism, if any. (1-2 sentences)

[2 marks]

3.b) Alice proposes that Paxos uses a fault-tolerant "total order broadcast" mechanism as part the communication in both of its two phases between a proposer and the acceptors. Discuss whether this is a valid approach to improve Paxos. (2-3 sentences)

[2 marks]

3.c) Explain whether a fault-tolerant consensus protocol can make progress when either the communicated messages or the processing at nodes are delayed arbitrarily. (3-4 sentences)

[3 marks]

3.d) What kind of message ordering is sufficient in a client-server system with multiple clients whose updates to the server state are commutative and the state update operation is idempotent ? *(1 sentence)*

[3 marks]

3.e) Two distributed processes S and R are connected through a network with lossy links that can drop, duplicate, or reorder messages, but the messages are guaranteed to be eventually delivered to R when they are re-delivered by S. Describe how to achieve a reliable communication abstraction using this network where a message is always received by R if it is sent by S. (2-3 sentences)

[2 marks]

3.f) Consider a collaborative calendar application implemented using a client-server model where multiple users (as clients) can simultaneously edit a calendar by either adding a new event or removing an existing event for a given date/hour. To edit the calendar, the clients use the add() and remove() remote procedure calls to add and remove events, respectively. A locking mechanism is used by the server to prevent conflicting edits to the calendar by multiple users at the same time. The clients can view a calendar by using a *read()* call, which returns a read-only copy of the calendar.

[Please turn over]

(i) As a new feature, the designers of the application consider an *offline editing mode* for devices to be able to edit a calendar while their devices are offline (e.g., due to poor network connection by mobile users). Discuss the trade-offs between the offline edit mode and the default read-only mode for offline clients. In your discussion refer to an impossibility theorem covered in the lectures. (At most 5-6 sentences)

[4 marks]

(ii) Discuss the repercussions of adding an offline editing mode to the calendar system in terms of consistency. *(3-4 sentences)*

[2 marks]

3.g) In a client-server system, one designated server T maintains accurate time using an expensive hardware (atomic) clock, while the other client nodes use cheap (quartz) clocks that are significantly less accurate. Initially, a human operator manually synchronises all the clocks of the clients with T's clock. However, eventually the clocks of the client nodes drift and are no longer synchronised with T. To synchronise their clocks with T, each client periodically communicates with T over the network and estimate the difference between its local clock and T's clock and then update their local clocks accordingly.

The communication between a client C and T is as follows: C sends a request to T when C's local clock is at t_1 and includes the timestamp t_1 in the request. T receives the request when its current clock is at t_2, and then sends a respond back to C when its clock is at t_3. The response message to C includes t_1, t_2, and t_3. C receives the response message from T when C's local clock is at t_4.

(i) Explain how the client C can estimate the time difference between its local clock and T's clock using the contents of the response message from T and t_4? *(3-4 sentences)*

[3 marks]

(ii) What assumption(s), if any, did you make in your calculation above? *(1-2 sentences)*

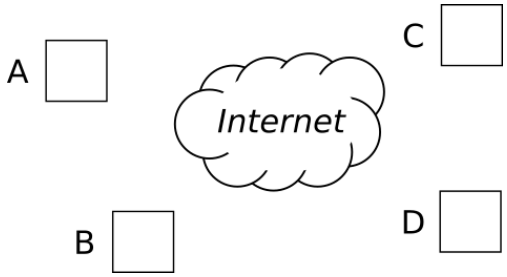
[2 marks]

Total 25 marks

[Please turn over]

Question 4

4.a Consider the below set of host computers, each of which is part of the same channel in a group communication service. We assume that UDP is used for the underlying message transport implementation, and that groups are operating on the public Internet.



(i) When sending a message with unreliable multicast semantics, how many network messages in total would be sent among the group? Explain your answer.

[3 marks]

(ii) When sending a message with reliable multicast semantics, using acknowledgements, how many network messages in total would be sent among the group? Explain your answer.

[3 marks]

(iii) For the reliable multicast case, how many we reduce the number of messages sent? Explain how your solution works, and how many network messages are now sent per group communication message. You should also explain the potential problems with your solution.

[4 marks]

(iv) When sending a message with atomic multicast semantics, how many network messages in total would be sent among the group? Explain your answer.

[3 marks]

4.b Using big-O notation, or an alternative precise description, how does network message volume increase with group membership size in each of the above schemes:

(i) Unreliable multicast:

[2 marks]

[Please turn over]

(ii) Reliable multicast using acknowledgements:

[2 marks]

(iii) Reliable multicast using your alternative approach:

[2 marks]

(iv) Atomic multicast:

[2 marks]

4.c Aiming to save on power and operational costs, an organisation wants to replace its centralised, single-server web cache system with an alternative peer-to-peer solution that uses a small amount of resource on each of its employees' personal computers.

Identify a peer-to-peer architecture which would be suitable for this objective, justify your choice, and explain how this approach might change the network characteristics of the organisation's internal LAN compared to the centralised single-server approach.

[4 marks]

Total 25 marks

--- End of Paper ---