Lancaster University

**COMPUTING AND COMMUNICATIONS – In-Person, Written Exam [2.5 hrs]**

**SCC.311      Distributed Systems**

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

**Use a _separate_ answer book for _each_ question.**

## Question 1

The vector clock algorithm is used in distributed systems to capture the partial ordering of events across multiple nodes. It allows nodes to maintain a logical timestamp called a vector clock, which consists of an array of integers. Each element in the array corresponds to a node in the distributed system. Here's the step-by-step algorithm for updating the vector clocks:

1. **Initialisation**: When a node $N_i$ starts, it initialises its local vector clock T to <0, 0, ..., 0>
2. **Event Occurrence**: Whenever an event occurs at node $N_i$ (e.g., sending or receiving a packet), it increments the element T[i] by 1. This represents the occurrence of a local event at node $N_i$.
3. **Sending a Message**: When node $N_i$ wants to send a message m, it sends the message along with its current vector clock (T) to the destination node via the network.
4. **Receiving a Message**: When node $N_i$ receives a message (m, T') from another node via the network, it updates its own vector clock according to the received vector clock (T') as follows. For each index j in the vector clock, $N_i$ takes the maximum value between its current timestamp T[j] and the corresponding timestamp T'[j] from the received vector clock. Additionally, $N_i$ increments the timestamp T[i] element by 1 to reflect the occurrence of the receive event.

**1.a.** Assign the vector clocks of each event (shown with a dot) and the clocks sent in the messages in the diagram below.
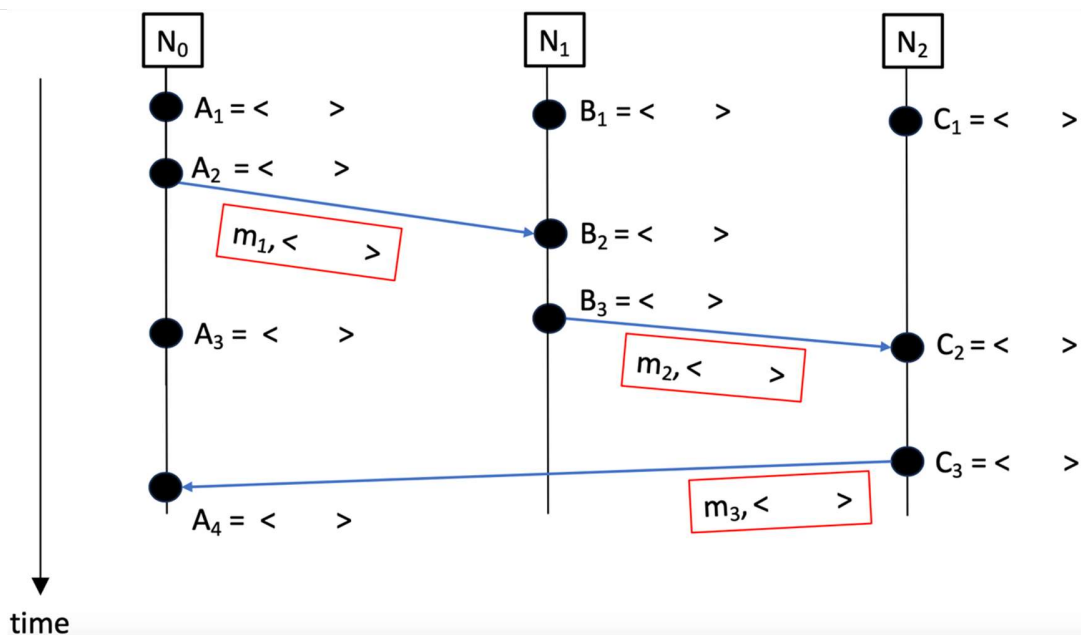


**[13 marks]**

**Question 1 continued.**

**1.b.** Based on the vector clocks of the events in the diagram, indicate which events happen **before** $C_2$ based on the happens-before relationship. (Note that this question is negatively marked for selecting incorrect options.)

**[3 marks]**

**1.c.** Based on the vector clocks of the events, indicate which events happen **after** the event $B_2$ based on the happens-before relationship (Note that this question is negatively marked for selecting incorrect options.)

**[2 marks]**

**1.d)** Write Java code for a function happensBefore() which takes two vector timestamps: $T_1$ and $T_2$ as arguments and returns True if $T_1$ happens-before $T_2$ and False otherwise. **Hint**: T1.length is the size of the array T1 in java.

```java
public class VectorTimestamp {

        public static boolean happensBefore(int[] T1, int[] T2)  {




















        }

}
```

**[4 marks]**

**1.e.** In terms of the happens-before relationship, what can you conclude about the order of the following pair of events? Justify your answer with the vector timestamp of each event.

      i.  $A_3$ and $C_2$
      ii. $C_1$ and $A_2$
      iii. $B_2$ and $A_4$

**[3 marks]**

**[Total 25 marks]**

**Question 2**

**2.a.** Which of the following statements are true concerning the detection of crash failures?
Note that points will be deducted for selecting incorrect choices, with the score not falling
below zero.

1. It is possible to determine whether a remote host has crashed accurately on the
   Internet.
2. Systems should be specifically designed to accommodate erroneous failure detections.
3. If a node has actually crashed, all nodes which interact with it will eventually be able to
   detect that crash.
4. A failure detector usually operates by sending a message to a remote node and waiting
   a fixed amount of time before reporting that node as having failed.

**[2 marks]**

**2.b.**
**i.** In a system designed to be tolerant of Byzantine failures, where there are three
malicious nodes, what is the minimum number of nodes in total that must be present
to detect the failure accurately?

**[2 marks]**

**ii.** How many communication rounds are needed within this group of nodes to detect
the problem using Lamport's algorithm?

**[2 marks]**

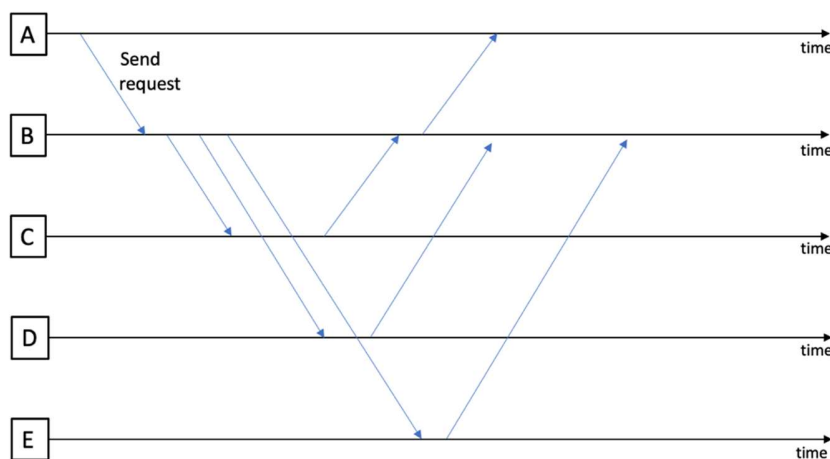**2.c.** The below diagram shows a form of service replication.



**Figure 1**

**Question 2 continues on the next page...**

i. State which kind of replication you think the diagram in Figure 1 represents.

**[2 marks]**

ii. What role does node A have in this system?

**[1 mark]**

iii. What role does node B have in this system?

**[1 mark]**

iv. Is the illustrated protocol a correct implementation of this kind of replication? Justify your answer.

**[5 mark]**

**2.d.**

i. A distributed hash table is a decentralised approach to store files. If every host issues one request for a random file every second, as we increase the number of hosts which are part of the network, but the total number of files being stored stays the same, how would you expect the volume of traffic from the perspective of an individual node to change? Explain your answer. Note: The number of files is much larger than the number of hosts.

**[2 marks]**

ii. In a client-server system to store files, a fixed number of hosts in the system are replica servers storing the files, while the rest are clients each requesting one random file from a randomly selected server. In this scenario, let's assume the number of clients increases with a fixed number of files; how would the volume of traffic change from the perspective of a server? Explain your answer. Note: The number of files is much larger than the number of clients.

**[2 marks]**

**Question 2 continued.**

**2.e.** Is it possible to implement a reliable failure detector using a reliable communication channel? Explain your answer.

**[3 marks]**

**2.f.** Discuss why indirect communication is appropriate for mobile environments where network coverage can be poor in some areas. *(2-3 sentences)*

**[2 marks]**

**2.g.** Name one way of achieving scalability in a distributed system.

**[1 mark]**

**[Total 25 marks]**

**Question 3**

Answer the multiple-choice questions below. Note that **points will be deducted for selecting incorrect choices**, with the individual score for each question not falling below zero. Unless otherwise stated, each question can have **multiple correct answers.**

**3.a.** Among the following use-cases below, which are potential use cases for a consensus protocol? Please select each option that applies.

a) Implementing state machine replication when there are no failures
b) Maximizing network throughput between a leader and backup replicas
c) Implementing FIFO group communication
d) Ensuring total order delivery of messages in a distributed system prone to failures
e) Ensuring consistency of state across replicas in the presence of failures

**[2 marks]**

**3.b.** Please select from the options below that apply to the Paxos consensus protocol.

a) It is a Peer-to-Peer approach
b) It involves a two-phase commit protocol
c) It provides Byzantine fault tolerance
d) It provides eventual consistency across replicas
e) It provides Crash fault tolerance

**[1 mark]**

**3.c.** What role does the leader node play in the Raft consensus protocol? Please select each option that applies.

a) Initiating elections
b) Proposing entries to be appended to the distributed log
c) Distributing votes to other nodes
d) Monitoring the health of other (i.e., follower) nodes
e) Receiving requests from clients, appending them to its log and sending them to followers

**[2 marks]**

**3.d.** What are the crucial requirements for achieving consensus in distributed systems? Please select each option that applies.

a) A quorum of replicas must reach an agreement
b) Complete absence of network partitions, meaning that any two replicas must be able to communicate at any given time
c) Asynchronous communication between replicas
d) All participating replicas must unanimously agree on a decision
e) Replicas must not crash due to a fault

**[2 marks]**

**3.e.** Which of the following protocols is **NOT** applicable to build a fault-tolerant, replicated state machine? Please select each option that applies.

a) Paxos
b) Raft
c) PBFT
d) Two-phase commit
e) Proof-of-Work

**[2 marks]**

**3.f.** Which of the following is a disadvantage of using traditional consensus protocols like Paxos or Raft in blockchain networks? Please select each option that applies.

a) Not Sybil-resistant
b) Low fault tolerance
c) Limited decentralisation
d) Not designed to tolerate Byzantine faults
e) Not designed to tolerate crash faults

**[2 marks]**

**3.g.** What is the primary purpose of a distributed log in consensus protocols like Raft or Paxos? Please select each option that applies.

a) Storing the state of each replica at the leader
b) Recording committed transactions for which consensus is reached
c) Caching frequently accessed transaction data
d) Maintaining network routing information
e) Recording client requests that are being replicated

**[2 marks]**

**3.h.** Which of the following statements are correct about Lamport's algorithm for logical clocks? Please select each option that applies.

a) Lamport clocks ensure that events are totally ordered.
b) Lamport clocks rely on global synchronisation for accuracy.
c) Lamport clocks can accurately capture potential causality among events.
d) Lamport clocks assign unique timestamps to each event.
e) Lamport clocks use vector of timestamps to each event

**[2 marks]**

**3.i.** Which of the following are challenges **NOT** addressed by PBFT? Please select each option that applies.

a) Crash fault tolerance
b) Byzantine fault tolerance
c) Tolerance to network partitions
d) Eventual consistency among replicas
e) State machine replication

**[1 mark]**


**3.j.** Which of the following are true when comparing PBFT (Practical Byzantine Fault Tolerance) to Paxos and Raft? Please select each option that applies.

a) PBFT  can tolerate a higher percentage of faulty replicas
b) PBFT requires fewer message exchanges to reach a consensus.
c) PBFT can tolerate crash faults
d) PBFT achieves lower latency in reaching consensus
e) PBFT provides tolerance to faults that are harder to detect

**[2 marks]**

**3.k.** Which of the following statements about active and passive replication? Please select each option that applies.

a) Active replication does not require a primary replica
b) Active replication is more suitable for a blockchain
c) Passive replication is more complex to implement than active replication
d) Passive replication involves replicas executing requests independently
e) Active replication is not tolerant to crash faults

**[2 marks]**

**3.l.** Which of the following best describes eventual consistency? Please select each option that applies.

a) All replicas are guaranteed to have the same state at all times
b) Consistency is achieved immediately after an update operation
c) Replicas may have temporarily divergent states but will eventually converge
d) Consistency is achieved through strict synchronisation of all replicas
e) Eventual consistency is not applicable in distributed systems

**[1 mark]**

**3.m)** Which of the following statements accurately describes the relationship between eventual consistency and conflict resolution in distributed systems? Please select each option that applies.

a) Eventual consistency guarantees conflict-free operation, eliminating the need for conflict-resolution mechanisms
b) Eventual consistency ensures that conflicts are immediately resolved to maintain consistency across replicas
c) Eventual consistency acknowledges the possibility of conflicts and provides mechanisms to detect and resolve them over time
d) Eventual consistency relies solely on strong consistency to prevent conflicts from occurring
e) Eventual consistency is incompatible with conflict resolution, as it prioritises availability over consistency

**[1 mark]**

**3.n)** Which of the following statements accurately describes the role of conflict-free replicated data types (CRDTs) in achieving eventual consistency and resolving conflicts in distributed systems? Please select each option that applies.

a) CRDTs guarantee immediate consistency across all replicas, eliminating the possibility of conflicts
b) CRDTs are only applicable in systems where strong consistency is prioritised over availability
c) CRDTs provide data structures and algorithms that ensure conflict resolution without requiring coordination among replicas
d) CRDTs rely on centralised authorities to resolve conflicts and maintain consistency
e) CRDTs are incompatible with eventual consistency, as they prioritise availability over consistency.

**[1 mark]**

**3.o)** Which statement accurately reflects the trade-offs described by the CAP theorem? Please select each option that applies.

a) In distributed systems, it is always possible to achieve both high availability and strong consistency simultaneously
b) The CAP theorem states that a distributed system can achieve only two out of three properties: consistency, availability, and partition tolerance
c) Consistency refers to the ability of a system to remain operational and responsive despite network partitions
d) Achieving strong consistency in a distributed system requires sacrificing partition tolerance
e) CAP theorem prioritises partition tolerance over both consistency and availability

**[2 marks]**

**[Total 25 marks]**

**Question 4**

Answer the multiple-choice questions below. Note that **points will be deducted for selecting incorrect choices**, with the individual score for each question not falling below zero. Unless otherwise stated, each question can have **multiple correct answers.**

**4.a.** Which of the following statements about Java RMI (Remote Method Invocation) is correct? Please select each option that applies.

a) Objects sent over the network using RMI must implement a remote interface
b) Java RMI allows Java objects to invoke methods only on objects located on the same JVM
c) Java RMI is suitable for client-server and not for peer-to-peer systems
d) Java RMI is not suitable to execute by a caller that resides in a different host than the remote object
e) Java RMI supports passing Java objects as arguments or return objects in remote method calls

**[2 marks]**

**4.b.** When a client invokes a method on a remote object located on a server in Java RMI, where does the execution of the method take place? Please select each option that applies.

a) The client first retrieves the object from the server and then executes the method locally
b) On the server where the remote object is located
c) Execution can happen on either the client side or the server side, depending on the implementation. d) Execution happens on both the client and server sides simultaneously
d) Execution is handled by the RMI registry.

**[2 marks]**

**4.c.** What is the primary difference between horizontal and vertical scaling? Please select each option that applies.

a) Horizontal scaling increases the size of individual resources, while vertical scaling adds more resources of the same type
b) Horizontal scaling adds more resources of the same type, while vertical scaling increases the size of individual resources
c) Horizontal scaling involves distributing workload across multiple servers, while vertical scaling involves upgrading a single server
d) Horizontal scaling is typically used for databases, while vertical scaling is used for web servers
e) Horizontal scaling is more cost-effective than vertical scaling

**[2 marks]**

**Question 4 continues on the next page…**

**4.d.** Which of the following statements about space and time uncoupling are NOT correct? Please select each option that applies.

a) Space and time uncoupling allows for messages to be produced and consumed asynchronously, decoupling the timing and location of message production and consumption
b) Distributed shared memory is an indirect communication approach with high scalability
c) Message queues are an example of space and time-coupled communication, where producers and consumers are closely synchronised in both space and time
d) Space and time uncoupling enables better scalability and fault tolerance in distributed systems
e) Space and time uncoupling eliminates the need for buffering messages in communication systems

**[2 marks]**

**4.e.** Which ordering semantics guarantees that events are delivered in the order they were sent by the sender? Please select each option that applies.

a) Global time ordering
b) FIFO ordering
c) Causal ordering
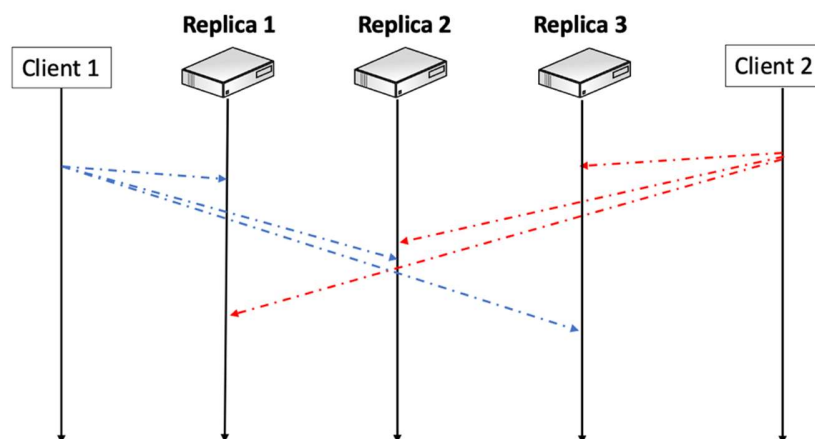d) Total ordering
e) Synchronous ordering

**[2 marks]**



**Figure 2**

**Question 4 continues on the next page…**

**4.f.** In Figure 2, two clients send messages to three replicas. At each replica, the incoming messages are delivered to the application in the order that they arrive. Indicate whether total ordering is achieved in the Figure above. Explain why or why not?

**[2 marks]**

**4.g.** Suppose that the two clients in Figure 2, stamp each message they send with their current system time. Also assume that the two clients have perfectly synchronised system clocks. Each replica temporarily buffers an incoming message, orders the messages by their timestamp, and then delivers them to the application. Indicate whether this approach can guarantee total ordering on the Internet. Explain why or why not.

**[2 marks]**

**4.h.** Which ordering semantics ensures that events are delivered in the order they occurred globally across the entire distributed system? Please select each option that applies.

a) Global time ordering
b) FIFO ordering
c) Causal ordering
d) Total ordering
e) Synchronous ordering

**[2 marks]**

**4.i.** In a distributed system with causal ordering, process A sends a message to processes B and C, and then C sends a message to processes A and B after receiving A's message. Please select each option that applies.

a) B must order A's message before process C
b) B must order C's message before A's message
c) B will order A's messages according to their order of arrival
d) B can order A's messages arbitrarily
e) A's message is related to C's message according to happens-before relationship

**[2 marks]**

**4.j.** Explain an approach that guarantees clients 1 and 2 in Figure 2 can individually achieve FIFO ordering. Describe what is required from both the client and the replicas.

**[2 marks]**

**4.k.** Bob proposes the following group communication mechanism where one node among N nodes is designated as the leader. Each node sends its messages to the leader, which then broadcasts them to all the nodes using a 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 mark]**

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

**[2 marks]**

    iii. Suggest one potential improvement or modification to Bob's broadcast mechanism to address its drawbacks.

**[2 marks]**

**[Total 25 Marks]**

**---End of Paper---**