Distributed Systems
Monsoon 2025
Mid Semester Examination
International Institute of Information Technology, Hyderabad
September 25, 2025. 10:00 AM to 11:30 AM

Instructions: Read the following points carefully.

1. The question paper has 10 questions spread across two sections for a total of 45 points.
2. The question paper has **TWO** pages. Ensure that your question paper is properly printed.
3. No clarifications shall be provided during the exam. If you need to make any assumptions to answer any question, state those assumptions explicitly. Answers arising out of incorrect or unsuitable assumptions may not get any points.
4. Answer all parts of each question contiguously.
5. Excessively verbose answers may attract negative points.

Section A -- Short Answer Questions. This section has five questions. Answer each question in no more than two to three paragraphs. (Points: 5x2 = 10)

1. Define the problem of termination detection and write in symbols the condition needed for termination. **(CO-1)**

2. Show an example of Maekawa's algorithm and how it can lead to deadlocks if the algorithm does not use Relinquish and Inquire messages. **(CO-1)**

3. Explain the purpose of the Combiner and Partitioner in the context of Map-Reduce. **(CO-2)**

4. Describe the differences between stateful and stateless protocols. What are some consequences of the differences mentioned in your answer. Illustrate with suitable examples. **(CO-3)**

5. Write a small note on inherent complexity and accidental complexity. What aspects of these do Map-Reduce and gRPC suffer from or support. **(CO-2)**

Section B. Long Answer Questions. This section has 5 questions. Answer each question in enough detail. Each question is for 7 points.

1. Consider the following definition of logical time. For a fixed integer k > 0, each process maintains a vector of size k, including its own local logical time. Processes piggyback their vector on messages they send. Processes update their vector based on their local logical time and the vector received on messages. Processes update their vector for local events too just like in scalar time. Answer the following questions based on the above description.

   a) Write the rules for updating the top-k logical time.
   b) Show an example with four processes and dozen events across the processes with k = 2.
   b) Is this logical time consistent? Justify your answer.
   c) Is this logical time strongly consistent. Justify your answer.

(Points: 2+1+2+2=7) **(CO-1)**

2. List the metadata that a GFS master needs to store with respect to each file. If the master in GFS has 128 MB space for storing this metadata, for how many files can the master keep information of? Discuss some advantages and disadvantages of keeping the metadata in RAM vs in disk. Make and state any assumptions you need to arrive at this answer.
(Points: 2+2+3=7) **(CO-3)**

3. Recall the algorithm of Lamport for guaranteeing mutual exclusion in a distributed system. Answer the following questions.

a) Show an example of the algorithm with four processors indicating the request/ reply/ and release messages.
b) What are the assumptions that the algorithm makes? Where is each such assumption needed in the algorithm? Explain your answer.
c) Suppose that a process P_i is executing in the critical section. Is the request of P_i at the top of the request queue at all sites when Pi entered the critical section? Justify your answer suitably.

(Points:2+2+3=7) **(CO-1)**

4. Define the notion of a consistent global state using both symbols and words. Define what is meant by causal order of message delivery with a suitable example. What are the typical challenges that algorithms for capturing a consistent global state must contend with. How does causal order of message delivery help design algorithms for capturing a consistent global state global state.
(Points: 1.5+1.5+1.5+2.5=7) **(CO-1)**

5. Recall the algorithm for termination detection using distributed snapshots. Answer the following questions.

(a) What assumptions does the algorithm make and why are these assumptions required for the algorithm?
(b) It is claimed that the last process to terminate will have the largest clock value. Justify this claim.
(c) What is the worst-case number of control messages triggered by the termination detection algorithm?
(d) Explain with a suitable example as to when this worst-case occurs.

(Points: 1.5+1.5+1+3=7) **(CO-1)**