Politecnico di Milano
090950 – Distributed Systems

Prof. G. Cugola – February 14th, 2024

Rules:

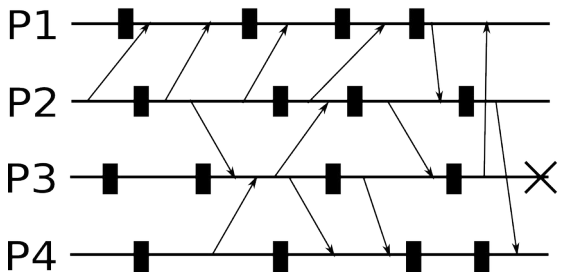
- You are **not allowed** to use books, notes, or other material.
- You can answer in Italian or English and use either pen or pencil.
- Total time for the test: 2 hours.

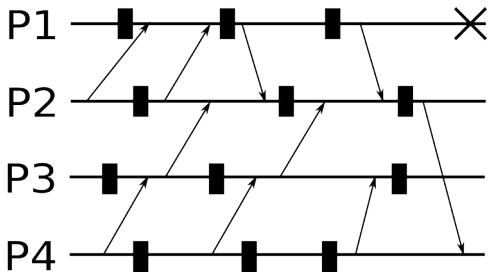
---

1. Describe mobile code paradigms making appropriate examples.

2. Describe and compare the various approaches to synchronize clocks in a distributed system. Now suppose you have to correlate readings of multiple, geographically distributed vibration sensors to determine the origin of an earthquake with a precision of less than 1km. Which clock synchronization approach would you use and why (seismic waves travel at a max speed of 10km per second)?

3. Calculate the recovery line for the two diagrams below using the rollback-dependency graph for the first one, the checkpoint dependency graph for the second one. **Finally, explain when this kind of graphs are used and how the data to build them is collected.**





4. Describe how to efficiently provide reliable group (multicast) communication among a set of reliable processes. How things change if processes may fail? Explain how the problem changes and how to address the new situation.

5. You are implementing an application where multiple processes access a replicated data store. The application requires FIFO consistency, but the data store you are using provides an eventual consistency model. Assuming there are no failures, does the data store satisfy the application's requirements? If not, can you implement a mechanism to ensure FIFO consistency on top of the eventually consistent data store? How?

6. Four peers (IDs = 2, 3, 4, 6) participate in a circular DHT with finger table using the CHORD protocol. Assume that the DHT uses 3-bits to represent the node IDs and the Keys. (a) Show the routing tables of the four peers. (b) Peer 3 wants to retrieve the value of an object having key 5. Show the exchange of messages required to search the desired value.

7. Consider the dataflow model for big data processing. (a) Describe the key characteristics of the model. (b) Describe what are the two architectures to implement it: scheduling of tasks and pipelining of tasks.