Question No 1 [20]

Consider an example of two program segments being executed by processor P0 and P1 as illustrated in the table below. The system consists of local memories (or caches) at processors P0 and P1, and a global memory. The three-state protocol assumed in this example corresponds to the state diagram illustrated below. Cache lines in this system can be either shared, invalid, or dirty. Each data item (variable) is assumed to be on a different cache line. Initially, the two variables `a` and `b` are tagged dirty, and the only copies of these variables exist in the global memory.

Fill in the table below by understanding the following steps. Use S for shared, I for invalid, and D for dirty.

1. Initially, the two variables `a` and `b` are tagged dirty, and the only copies of these variables exist in the lobal memory. This is done for you in the table as an example.
2. Processor 0 executes `read a`. Update the table accordingly.
3. Processor 1 executes `read b`. Update the table accordingly.
4. Processor 0 executes `a = a + n`, where `n` is the last digit of your roll number. Update
5. the table accordingly.
6. Processor 1 executes `b = b + n`, where `n` is the last digit of your roll number. Update
7. the table accordingly.
8. Processor 0 executes `read b`. Update the table accordingly.
9. Processor 1 executes `read a`. Update the table accordingly.
10. Processor 0 executes `a = a + b`. Update the table accordingly.
11. Processor 1 executes `b = a + b`. Update the table accordingly.

**Example**

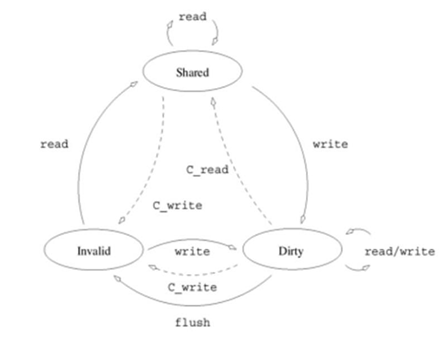
If a student's roll number is 21F-9624, the last digit is 4. Therefore, the student will replace `n` with 4 in the operations:

Step 4: `a = a + 4`

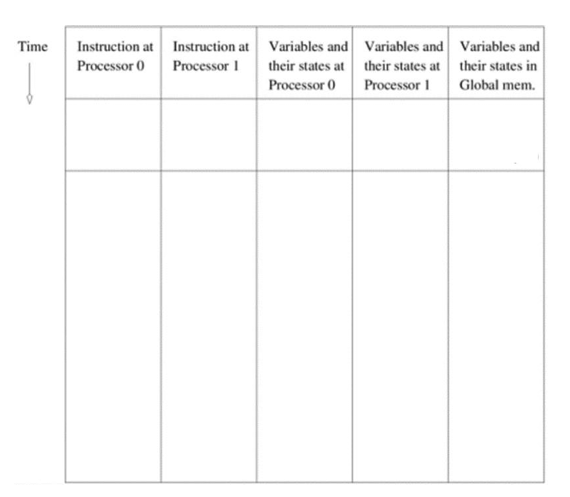
Step 5: `b = b + 4`

**Note:**

Students must follow their own roll number to determine the value of `n` (the last digit of their roll number) and use it in the operations. They should then track the state of variables `a` and `b` in the caches of P0, P1, and global memory after each operation.



**MAKE TABLE LIKE THIS:**



Question No 2 [10]

* Explain how the bisection bandwidth of an interconnect impacts scalability in parallel systems.
* A torus interconnect with 1024 nodes suffers from congestion due to hotspots in communication patterns.
* Compute the network diameter assuming a 2D torus topology.
* If message latency grows quadratically with hop count, derive a formula for expected delay as a function of processor count (P).
* Propose a congestion-aware routing algorithm to reduce network contention and improve throughput.

Question No 3 [20]

A space agency is designing a communication network for a Mars research station. The station consists of multiple research modules, each collecting and transmitting scientific data to Earth via a central relay station.

**Challenges:**

* Due to long distances, signal delays are unavoidable.
* Data needs to be aggregated from multiple modules and efficiently transmitted.
* Some modules may fail due to harsh environmental conditions.

**Discussion Questions:**

Interconnect Design: Would a bus-based, tree-based, or star-connected topology be most suitable for communication between the research modules and the relay station? Why?

Delay Handling: How should the topology be designed to minimize the impact of latency in data transmission?

Robustness: If one research module fails, which topology ensures that other modules can still communicate with the relay station?

Energy Efficiency: Given limited power supply on Mars, which topology minimizes energy consumption for data transfer?