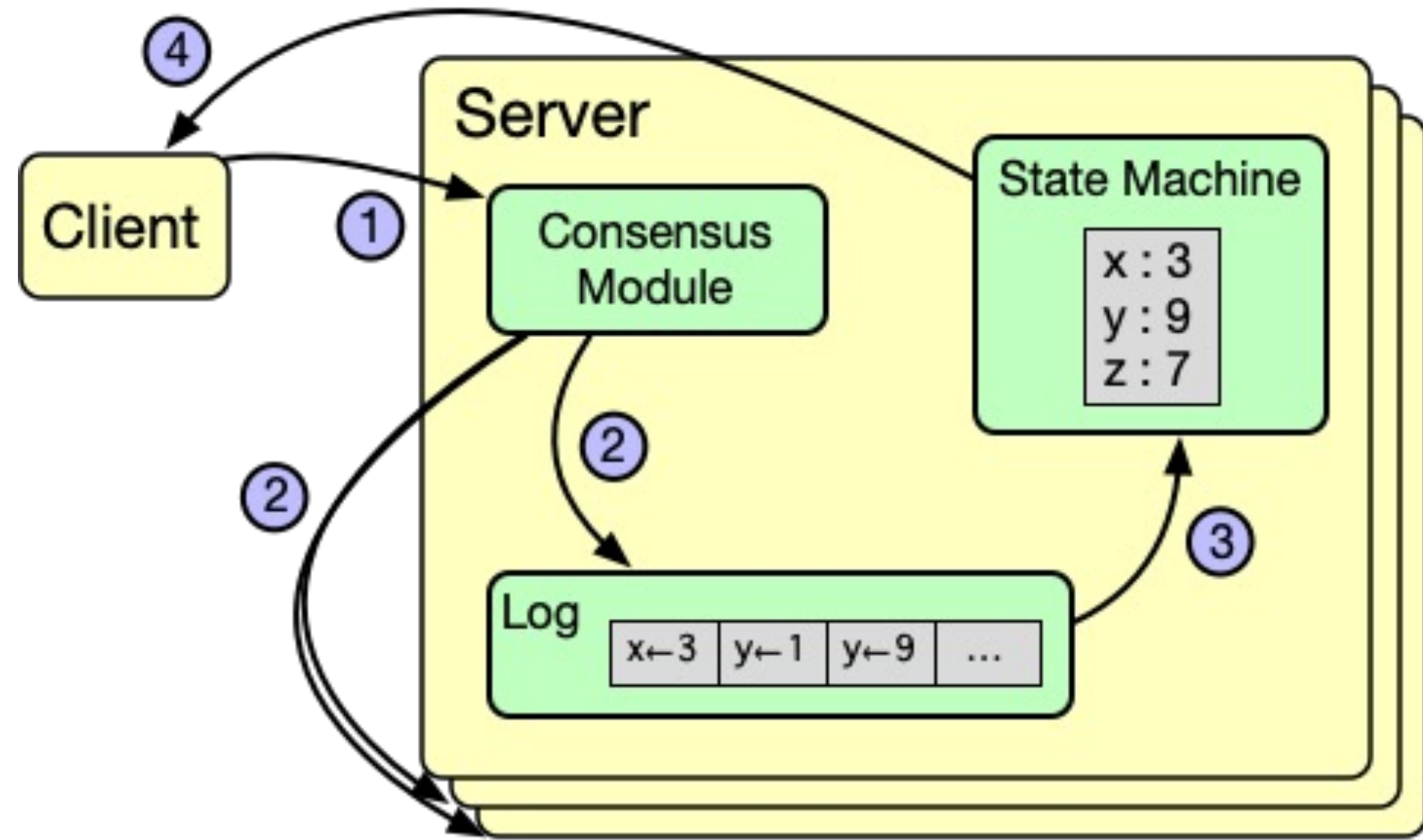


SSR: Disaggregated and Timing-Aligned Synchronous State Machine Replication for Modern Datacenter

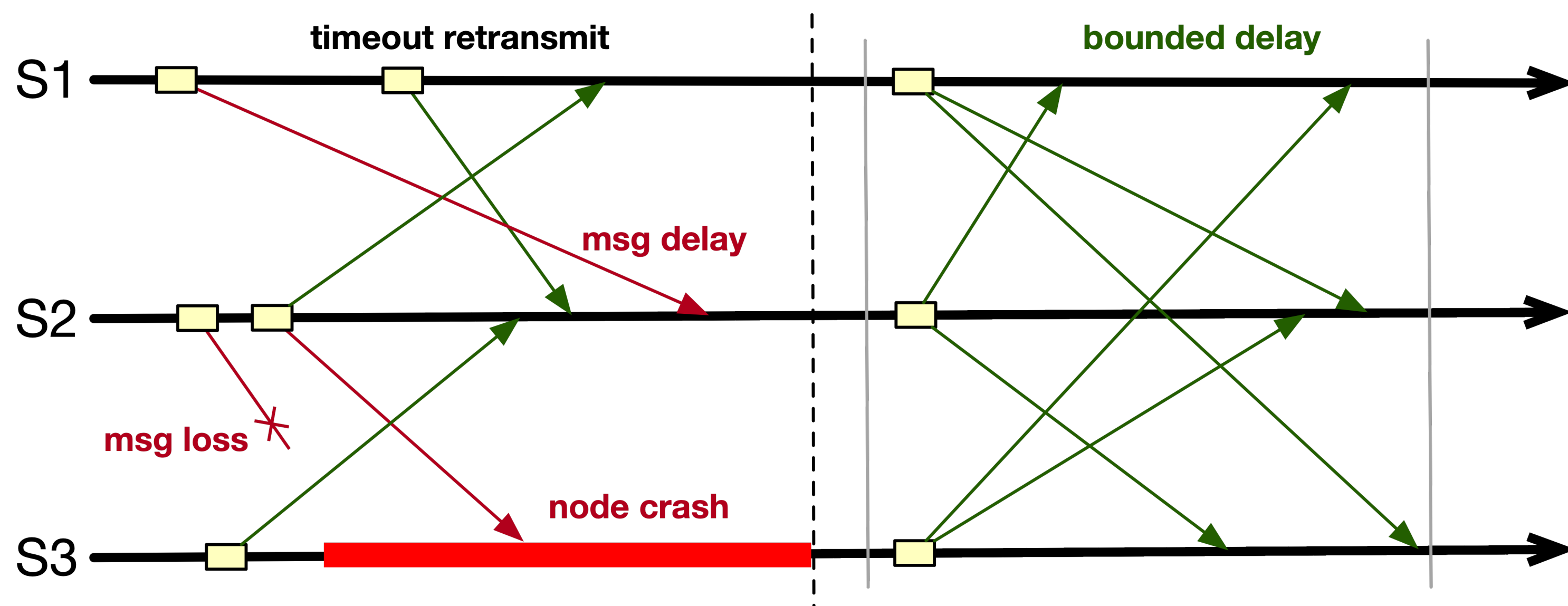
Yuke Ma[♠], Matthew Yao[♥], Paolo Costa[♣], Yiting Xia[♠]
Max Planck Institute for Informatics[♠], Microsoft Research[♣], Purdue University[♥]

State Machine Replication



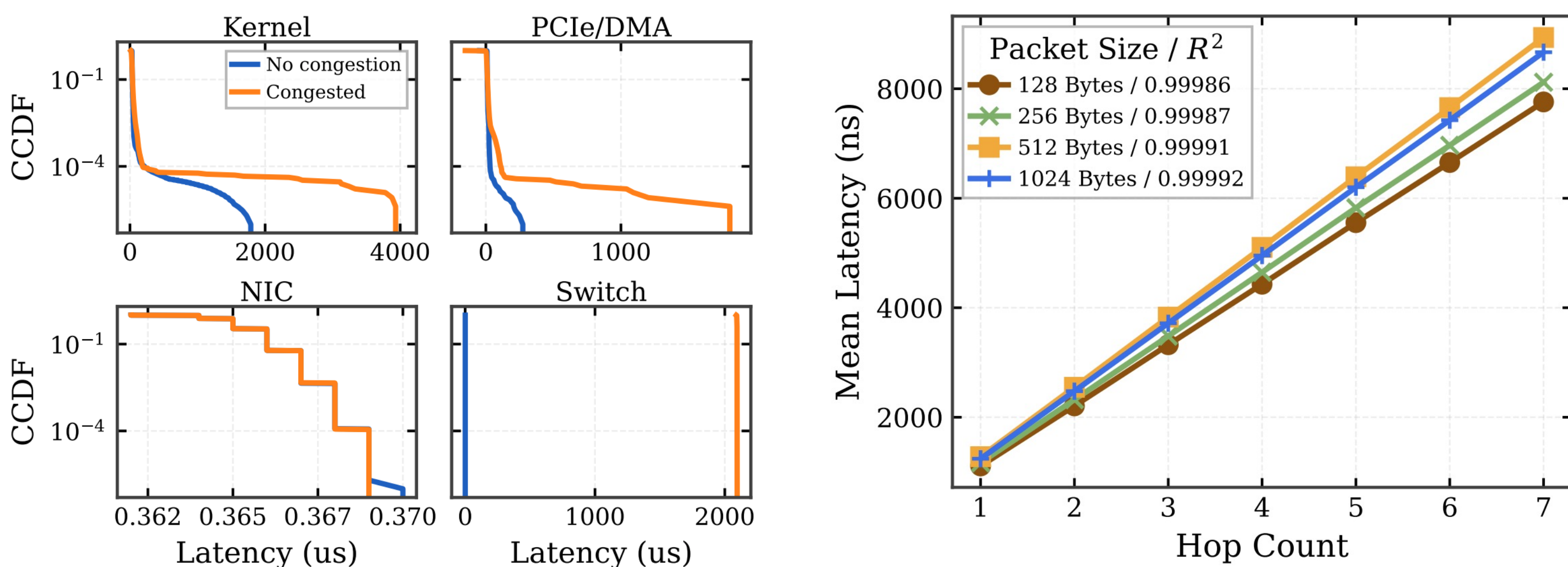
- distributed databases
- distributed key-value stores
- metadata services
- distributed lock services
- financial systems

Asynchronous vs. Synchronous System



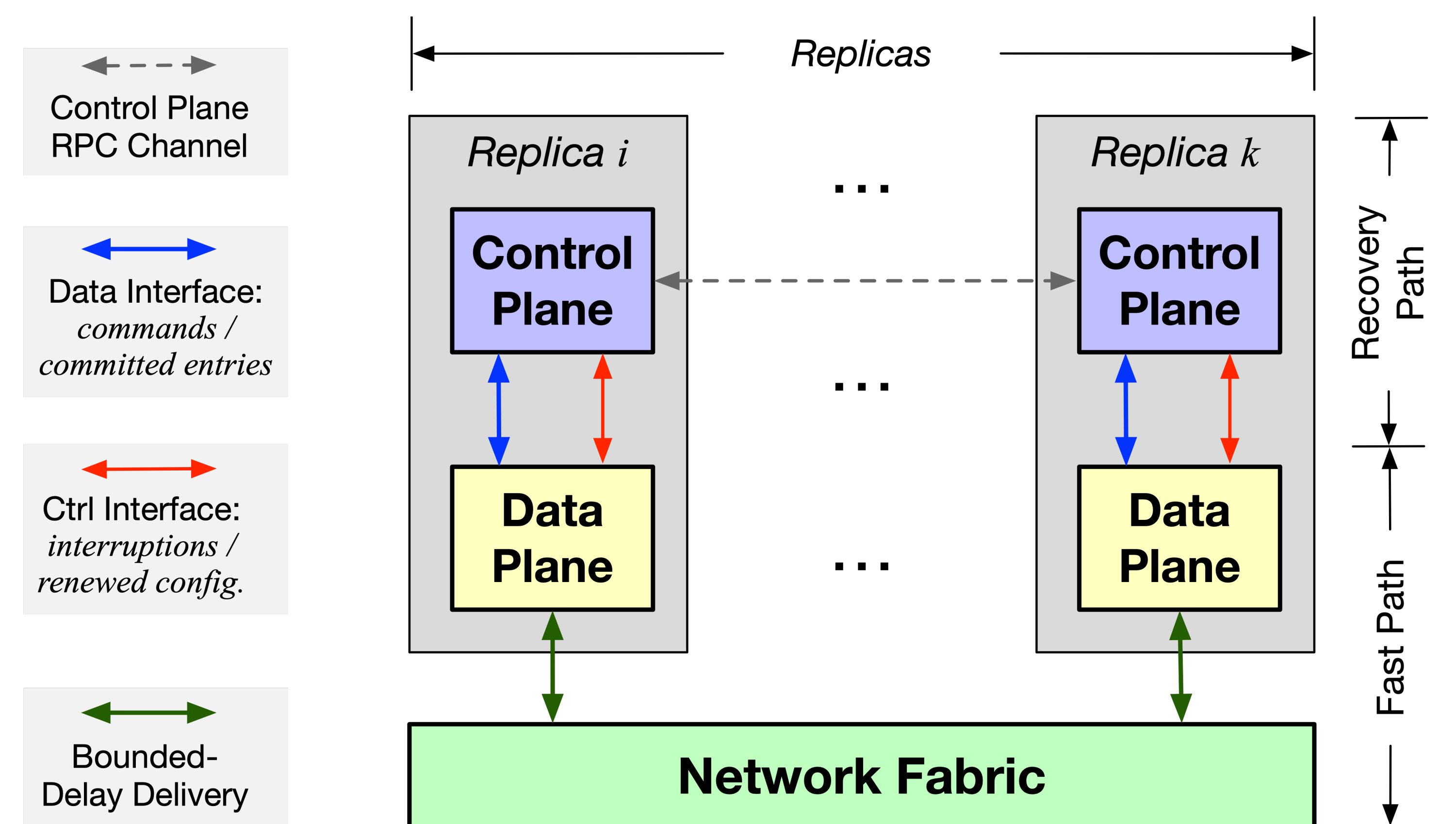
	Asynchronous	Synchronous
Message Delay	No upper bound	Known upper bound
Processing Time	No upper bound	Known upper bound
Clock Drift	No assumptions	Bounded

Split Timing model in Modern Datacenter



Host (PCIe/DMA, OS, Application) : Variant and long tail latency
Network (NIC, Switch) : Deterministic latency without queueing

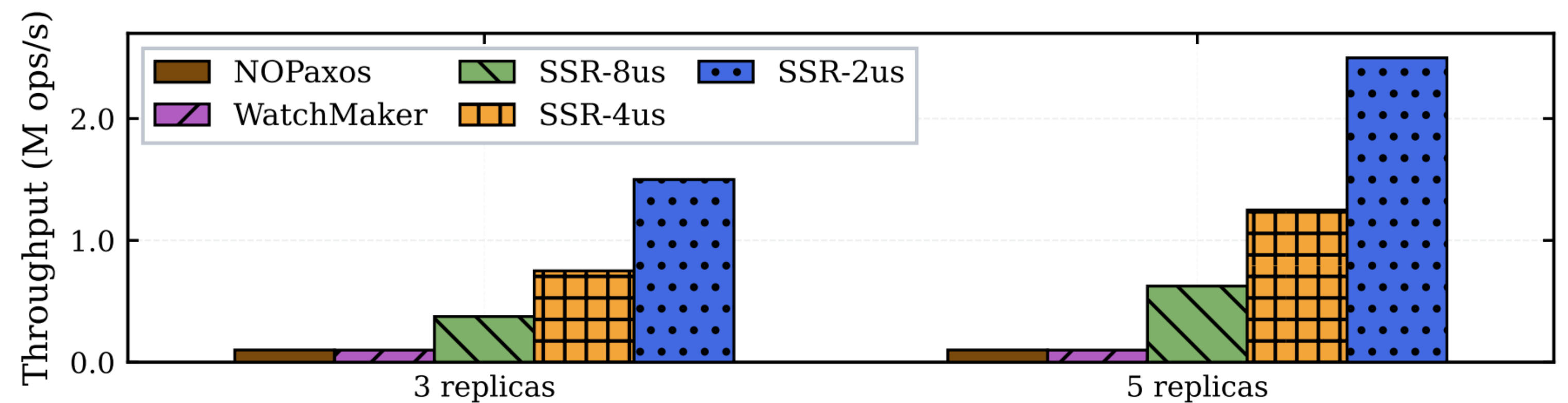
SSR Timing-Aligned Architecture



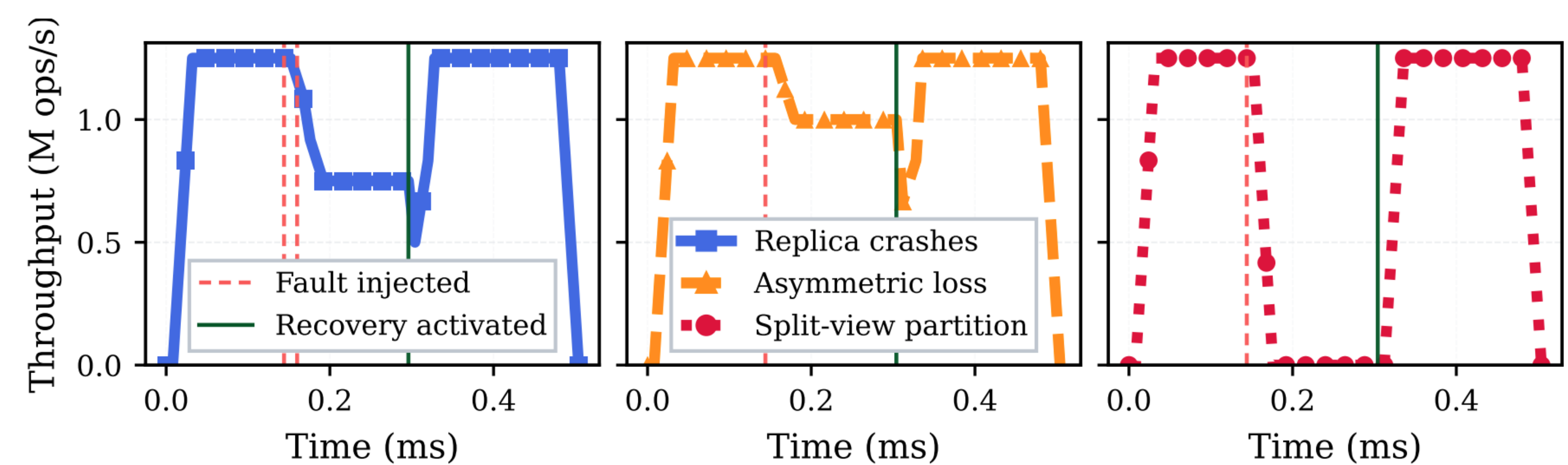
Control Plane: Application interaction, cluster management, failure recovery, and reconfiguration

Data Plane : Repetitive, and latency-critical consensus transactions

Preliminary Simulation Results



Throughput: NOPaxos and WatchMaker achieve 100 and 99 Kops/s. SSR reaches 1.49 Mops/s with 3 replicas and 2.49 Mops/s with 5



Failure recovery : Synchrony exposes failures within bounded rounds, enabling prompt recovery.

Latency: Consensus transactions complete in 2 bounded rounds

Two Subtle Examples of SSR's Failure Detection and Fail-stop Mechanism

