Next Generation State-Machine Replication Protocols Among Data Centers

Presented By

NEHME Mohamad-Jaafar

ERODS Team

Directed By

QUEMA Vivien

Co-Directed By

BEYDOUN Kamal and PALIX Nicolas

Outline

- State Machine Replication
- Total Order Broadcast
- Protocol Characteristics
- State Of The Art
- My PhD Goal

Outline

- **State Machine Replication**
- Total Order Broadcast
- Protocol Characteristics
- State Of The Art
- My PhD Goal

State-Machine Replication (SMR)

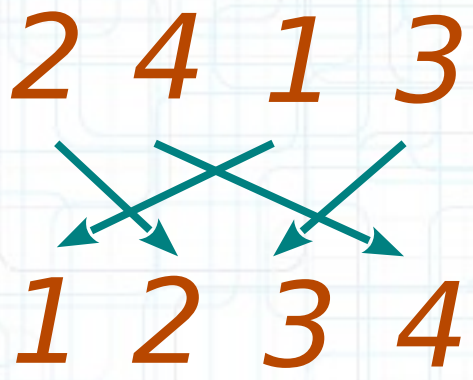
- Need for SMR
  - Storing data on replicas hosted on Data Centers (DCs)
- Fault Tolerance
- High availability
  - Being responsive to a lot of clients
- Strong consistency
  - Deterministic and atomic execution of commands
  - Same order of commands (Write/Read)

Outline

- State Machine Replication
- **Total Order Broadcast**
- Protocol Characteristics
- State Of The Art
- My PhD Goal

Total Order Broadcast (TOB)

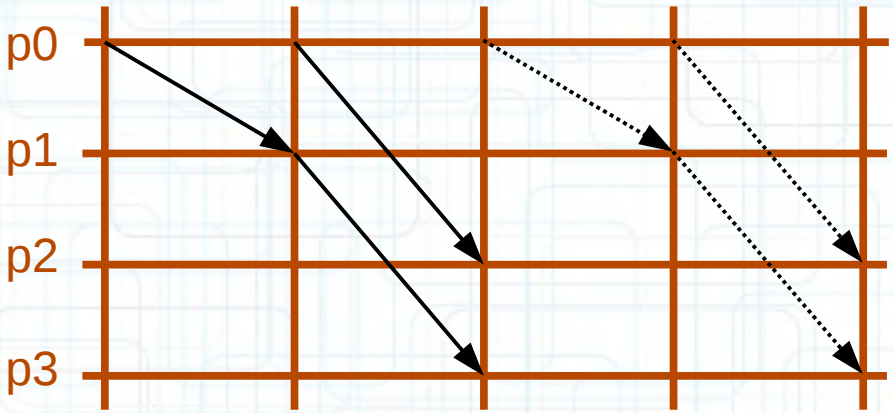
- Ensuring strong consistency
  - Broadcast and execute in order
- Correct/Faulty Processes
  - Crash, omission, timing or Byzantine failures [Classification]
- Uniform TOB
  - Uniform agreement
    - r delivers m → correct rs deliver m
  - Strong uniform total order
    - r delivers m, m' → rs deliver m,m'
- Propositions in the literature



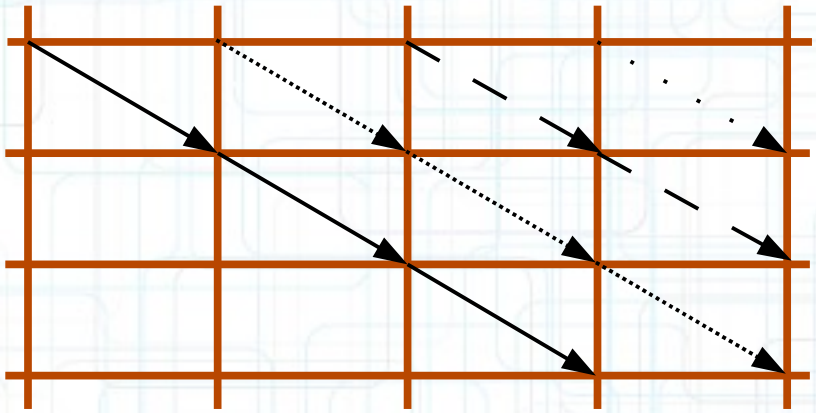
Outline

- State Machine Replication
- Total Order Broadcast
- **Protocol Characteristics**
- State Of The Art
- My PhD Goal

Protocol Characteristics



Latency = 2
Throughput = 1/2



Latency = 3
Throughput = 1

- Latency
  - Time to broadcast a message
- Throughput
  - Number of messages / time unit
- Comparison of two algorithms [LCR]
- Scalability

Outline

- State Machine Replication
- Total Order Broadcast
- Protocol Characteristics
- **State Of The Art**
- My PhD Goal

State Of The Art

- Classification by Défago *et al.*
  - Fixed-sequencer
  - Moving sequencer
  - Privilege-based
  - Communication history
  - Destination agreement
- LCR
- FastCast
- S-SMR
- P-SMR
- Clock-RSM

Protocol	Latency	Throughput	Scalable on DCs
LCR	Medium	Optimal	No
FastCast	Low	Optimal	No
S-SMR	High	Medium	Possible
P-SMR	Medium	Medium	Possible
Clock-RSM	High	Low	Yes

Outline

- State Machine Replication
- Total Order Broadcast
- Protocol Characteristics
- State Of The Art
- **My PhD Goal**

My PhD Goal

Can we scale with multiple DCs preserving low latency and optimal throughput?

- Problems to address
  - Sharing network infrastructure at WAN/LAN level
  - Scale with multi-core systems
- Research aim
  - New design(s) for UTOB protocols
  - Optimizing performance among DCs

References

[Class.] Xavier Défago, , André Schiper, and Peter Urban. “Comparative performance analysis of ordering strategies in atomic broadcast algorithms”. IEICE Trans. 2003 Inf. Syst. E86-D, 12, 2698–2709

[LCR] Rachid. Guerraoui, Ron R. Levy, Bastian Pochon, and Vivien Quéma, “Throughput optimal total order broadcast for cluster environments”. ACM Trans. Comput. Syst., vol. 28, no. 2, pp. 5:1–5:32, Jul. 2010.

[FastCast] Gautier Berthou and Vivien Quéma. “FastCast: a Throughput- and Latency-efficient Total Order Broadcast Protocol”. Middleware 2013 Volume 8275, 2013, pp 1-20

[S-SMR] Carlos Eduardo Benevides Bezerra, Fernando Pedone, Robbert van Renesse: “Scalable State-Machine Replication”. DSN 2014: 331-342

[P-SMR] Parisa Jalili Marandi, Carlos Eduardo Benevides Bezerra, Fernando Pedone: “Rethinking State-Machine Replication for Parallelism”. ICDCS 2014: 368-377

[Clock-RSM] Jiaqing Du, Daniele Sciascia, Sameh Elnikety, Willy Zwaenepoel, Fernando Pedone: “Clock-RSM: Low-Latency Inter-datacenter State Machine Replication Using Loosely Synchronized Physical Clocks.” DSN 2014: 343-354