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 \( r_s \) deliver \( m \)
  - Strong uniform total order
    - \( r \) delivers \( m, m' \) → \( r_s \) 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**
  - Time to broadcast a message
- **Throughput**
  - Number of messages / time unit
- **Comparison of two algorithms [LCR]**
- **Scalability**

Latency = 2
Throughput = 1/2

Latency = 3
Throughput = 1
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

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Latency</th>
<th>Throughput</th>
<th>Scalable on DCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCR</td>
<td>Medium</td>
<td>Optimal</td>
<td>No</td>
</tr>
<tr>
<td>FastCast</td>
<td>Low</td>
<td>Optimal</td>
<td>No</td>
</tr>
<tr>
<td>S-SMR</td>
<td>High</td>
<td>Medium</td>
<td>Possible</td>
</tr>
<tr>
<td>P-SMR</td>
<td>Medium</td>
<td>Medium</td>
<td>Possible</td>
</tr>
<tr>
<td>Clock-RSM</td>
<td>High</td>
<td>Low</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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


