Università della Svizzera italiana

Faculty of Informatics

Unicast Multi-Ring Paxos

Samuel Benz
Prof. Fernando Pedone

Introduction

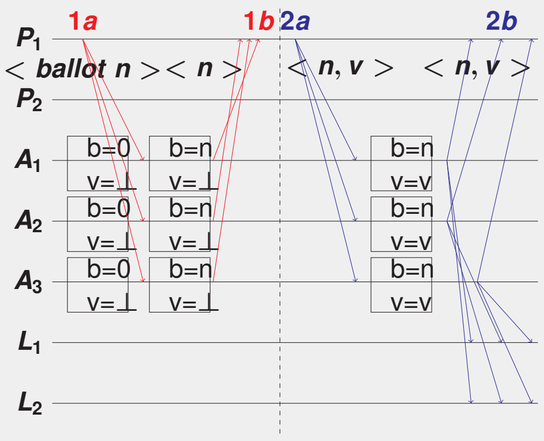
The rise of internet-scale services demands large distributed systems. Such huge infrastructures have to be coordinated and, to provide fault-tolerance, replicated. These aspects can be encapsulated in an atomic multicast protocol. Our approach to scaling an atomic multicast protocol is Multi-Ring Paxos. It uses multiple efficient rings of Paxos to provide scalable high throughput of atomically ordered traffic. The provided implementation contains all components needed to tolerate adverse conditions, such as message losses and process crashes.

Consensus and Atomic Multicast

1. **Termination:** Every correct process eventually decides.
2. **Agreement:** No two correct processes decide differently.
3. **Uniform integrity:** Every process decides at most once.
4. **Uniform validity:** If a process decides ***v***, then ***v*** was proposed by some process.
5. **Total order:** If two correct processes ***p*** and ***q*** deliver two messages ***m*** and ***m′***, then ***p*** delivers ***m*** before ***m′*** if and only if ***q*** delivers ***m*** before ***m′***.

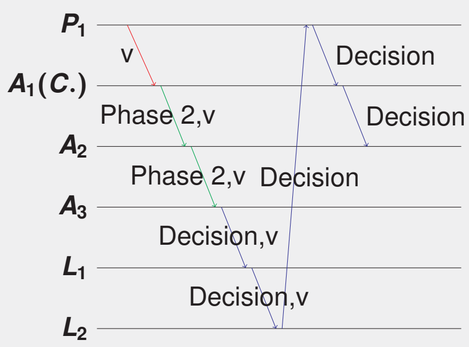
[Chandra *et al.* Unreliable failure detectors for reliable distributed systems. 1996.]

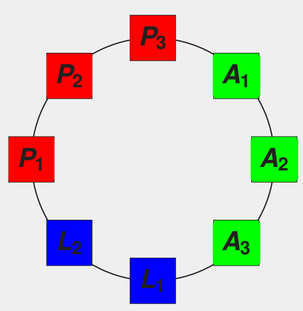
Paxos



[Lamport. The part-time parliament. 1998.]

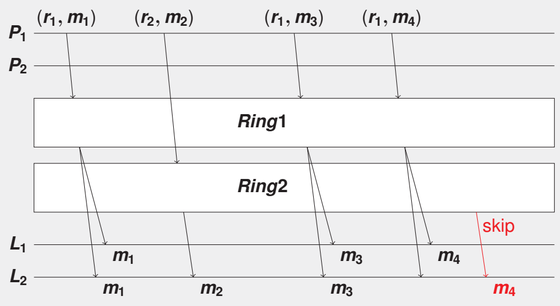
Ring Paxos

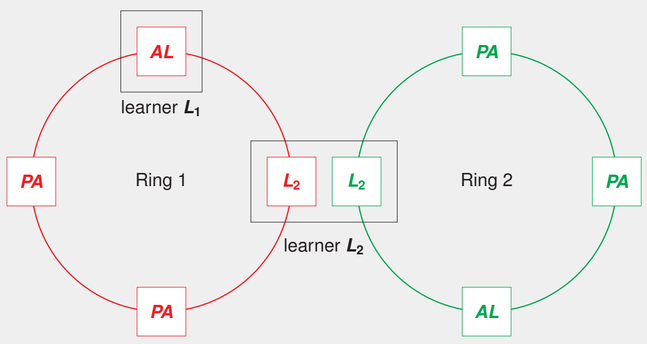




[Marandi *et al.* Ring paxos: A high-throughput atomic broadcast protocol. 2010.]

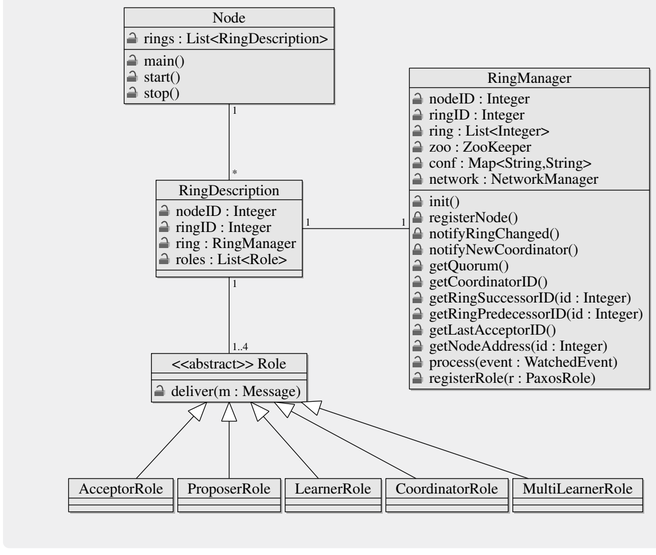
Multi-Ring Paxos





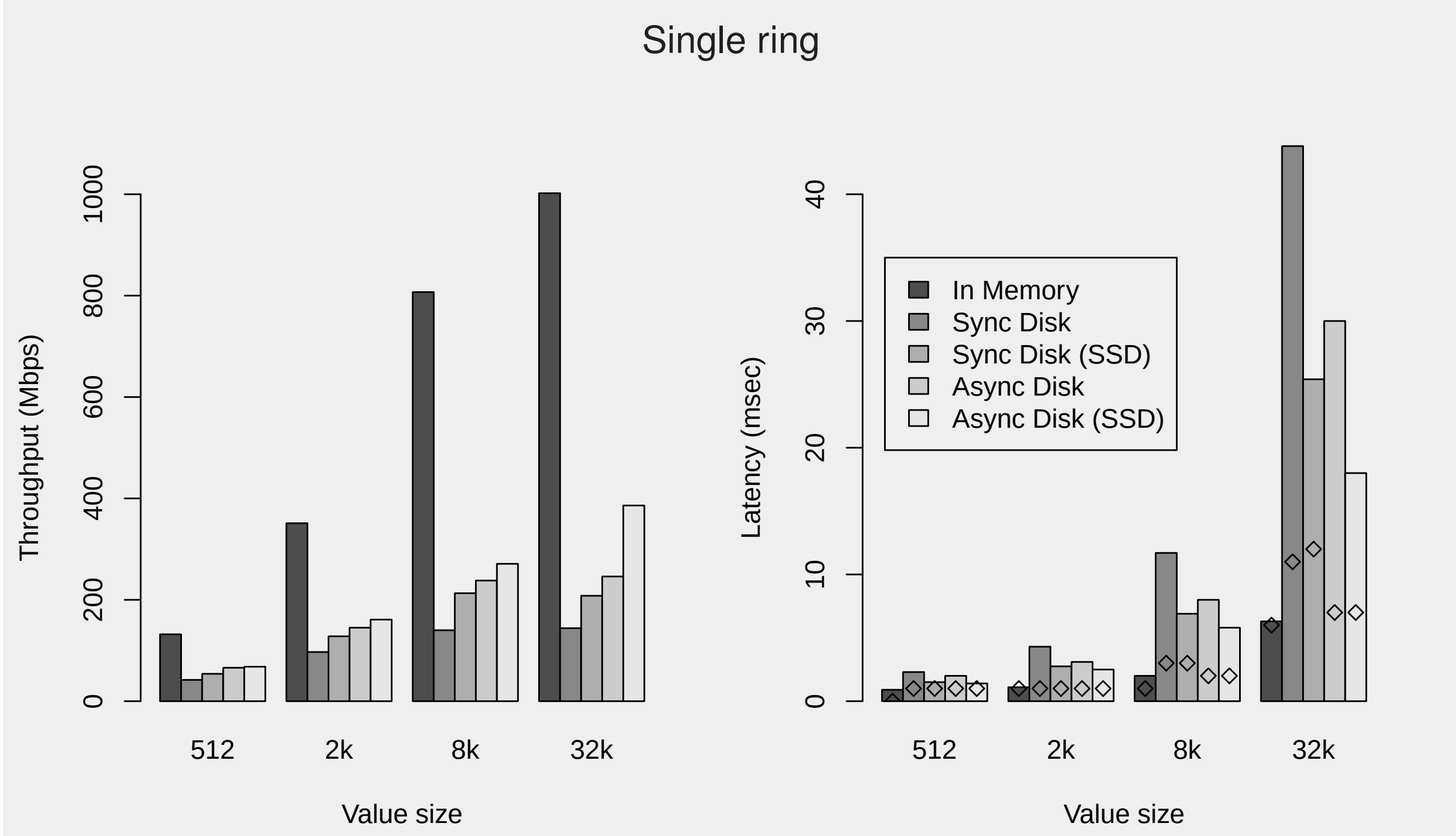
[Marandi *et al.* Multi-ring paxos. 2012.]

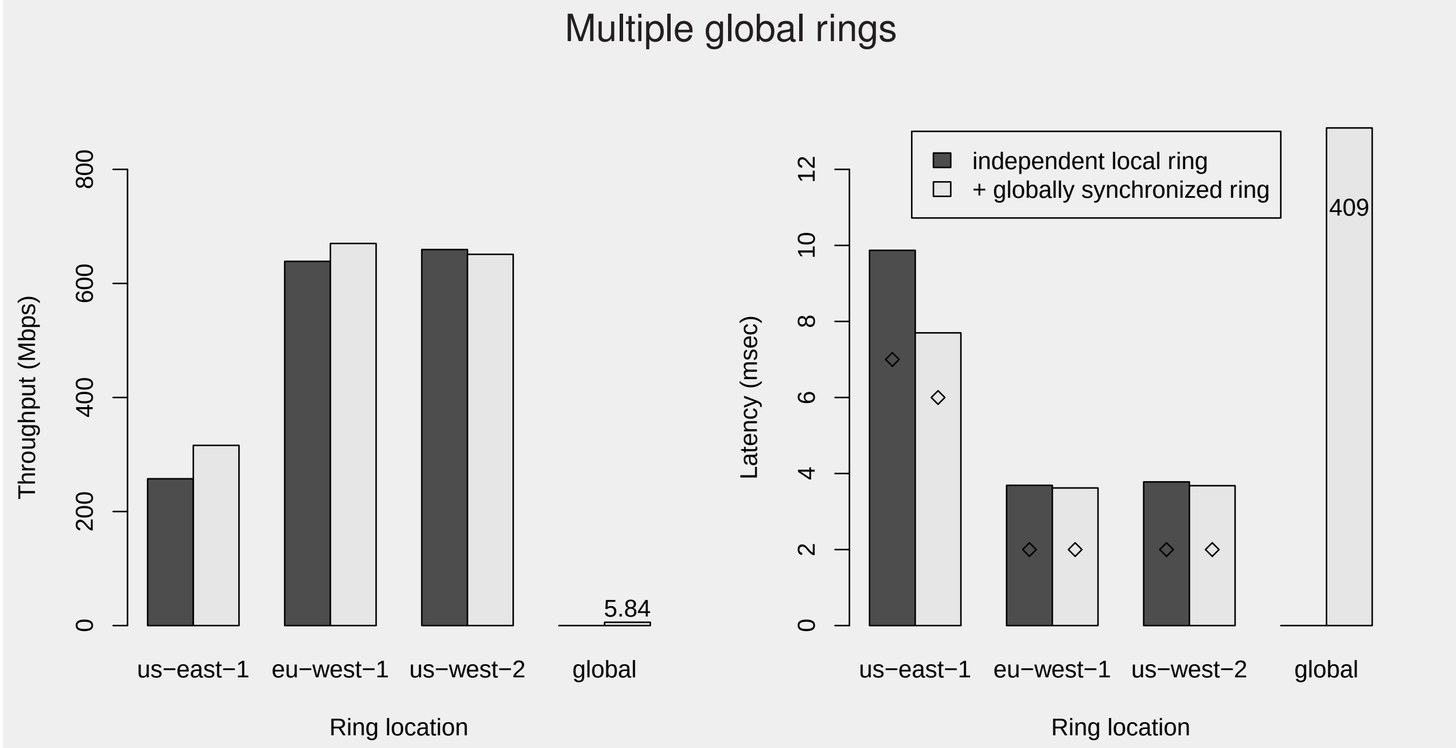
Implementation



https://github.com/sambenz/URingPaxos

Performance





Contributions

1. Complete implementation (”production level”) of an Atomic Multicast protocol
2. Detailed performance evaluation of Ring Paxos with different storage guarantees
3. Detailed performance evaluation of Multi-Ring Paxos in a globally distributed environment
4. Release of the implementation as open source under GPL (and already in use by groups at other universities)