Fleeing Paxos by Raft

December 1, 2017

Motivation

“

a pair consisting of an integer and a priest, using a lexicographical ordering, where

$$(13, \Gamma\rho\alpha\breve{\iota}) < (13, \Lambda\iota\nu\sigma\breve{\iota}) < (15, \Gamma\rho\alpha\breve{\iota})$$

...

To maintain B2, a ballot’s quorum was chosen to contain a μαδζδωριτ˘ιστ of priests. Initially, μαδζδωριτ˘ιστ just meant a simple majority. Later, it was observed that fat

”

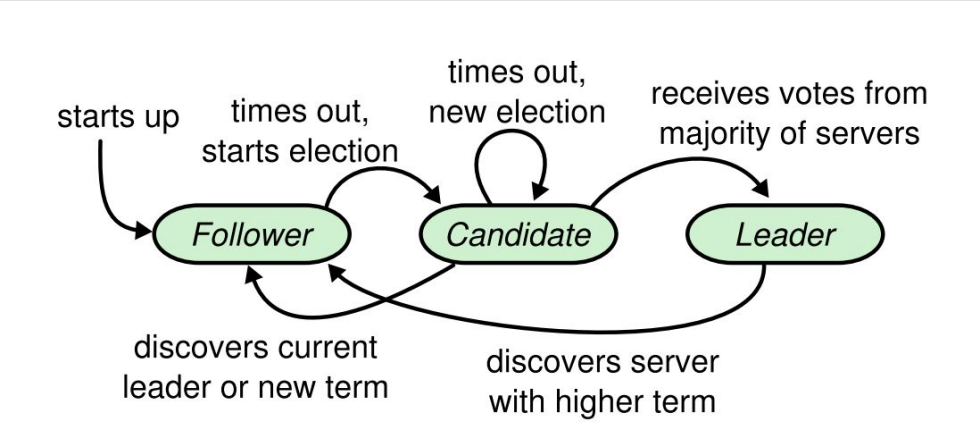
priests were less mobile and spent more time in the Chamber than thin ones,

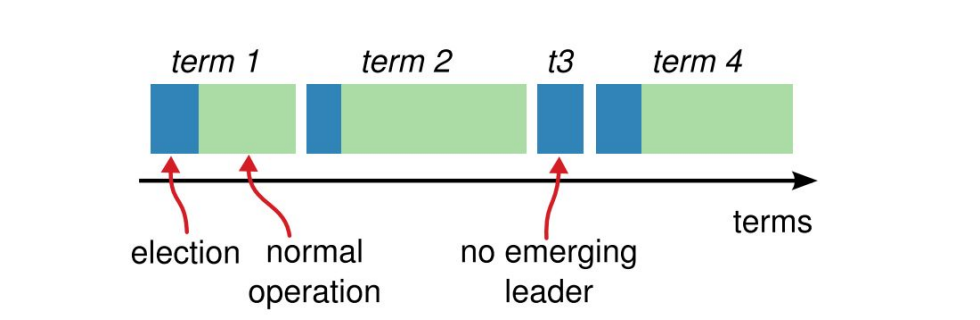
Raft is Designed for Understandability

Three largely independent subproblems:

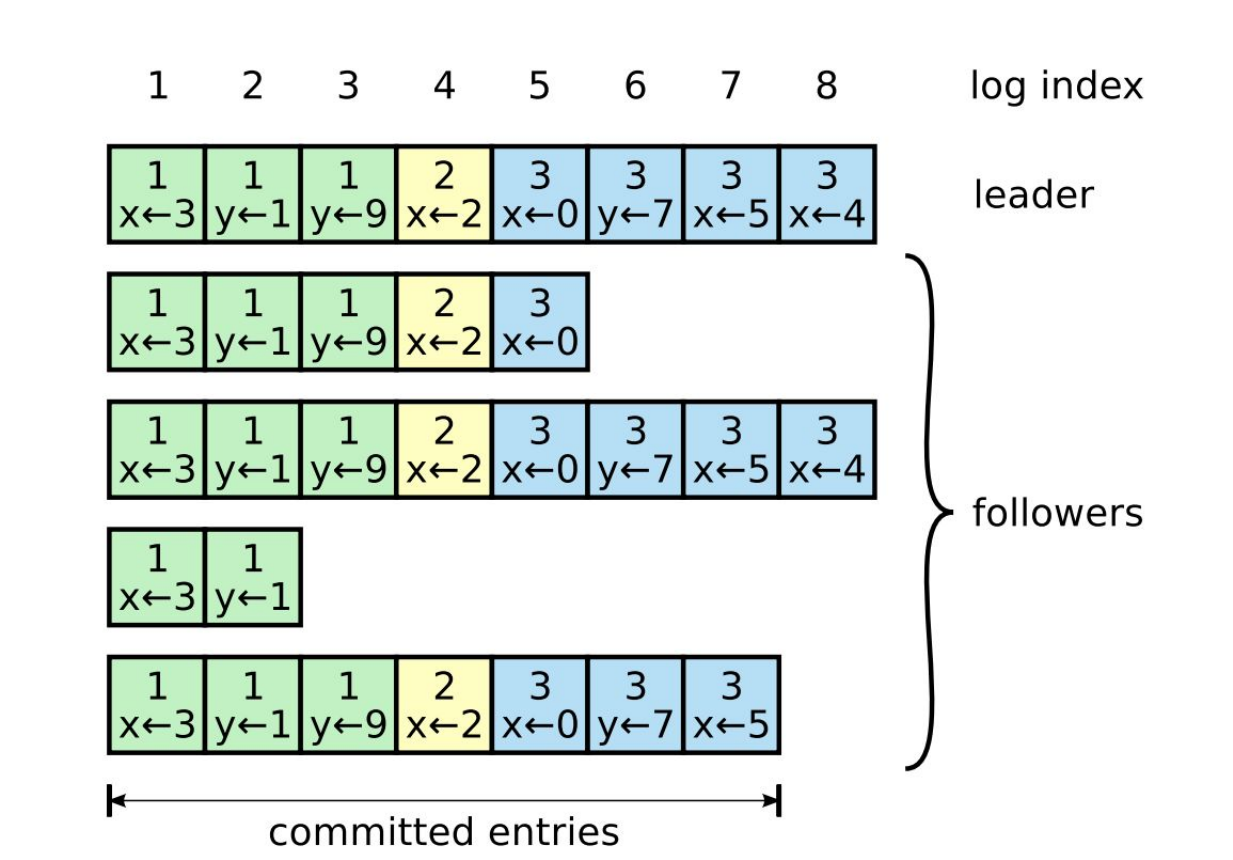
1. Leader Election
2. Log Replication
3. Safety

Structures & States





Structures & States (ctd.)



Subproblems

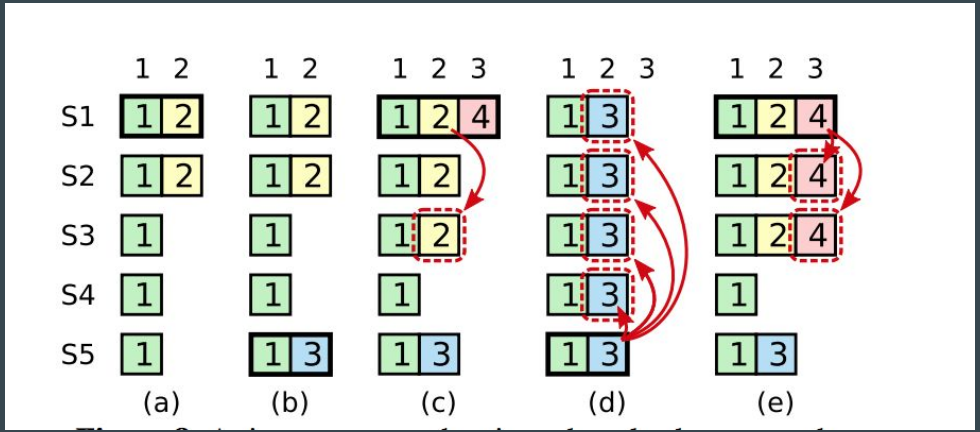
1. Leader Election
    a. Leader sends heartbeat to followers using empty AppendEntries RPC
    b. Followers each maintain their own randomized countdown timer to declare current leader dead
    c. If timer expires without a heartbeat, Follower increments term, becomes Candidate and sends RequestVote RPC to all other members
    d. If 50% of members reply with their vote, Candidate becomes new Leader

Subproblems

2. Log Replication
    a. All log updates sent by Leader
    b. Same log index && same term value => same log entry
    c. Previous log index and term sent with each update to ensure previous entry is correct
    d. If the previous entry is correct before next entry applied, then all earlier entries are also correct

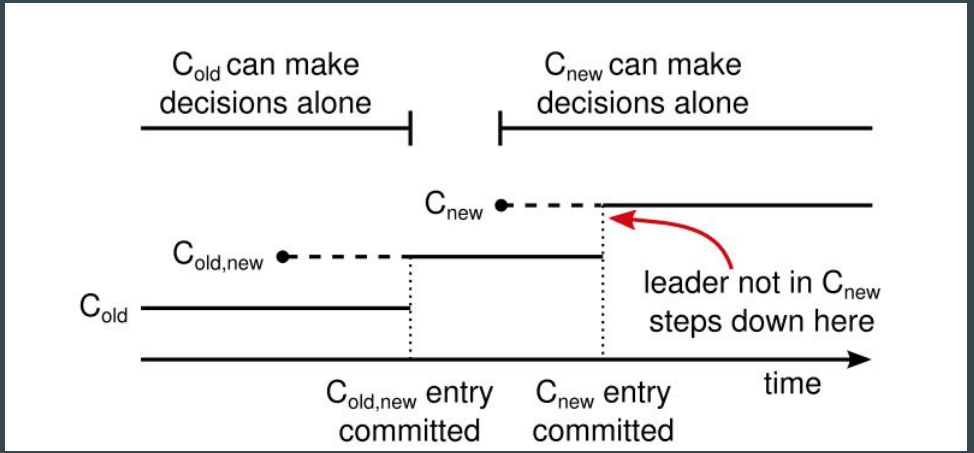
Subproblems

3. Safety
    a. Only candidates with all log entries committed can be elected leader
    b. Before granting a vote to a Candidate, the voter checks the Candidate's log for completeness.
    c. A Candidate with an incomplete log will not be voted in as Leader.
    d. New leaders will not commit previous term's entries using replica counting



Batteries Included

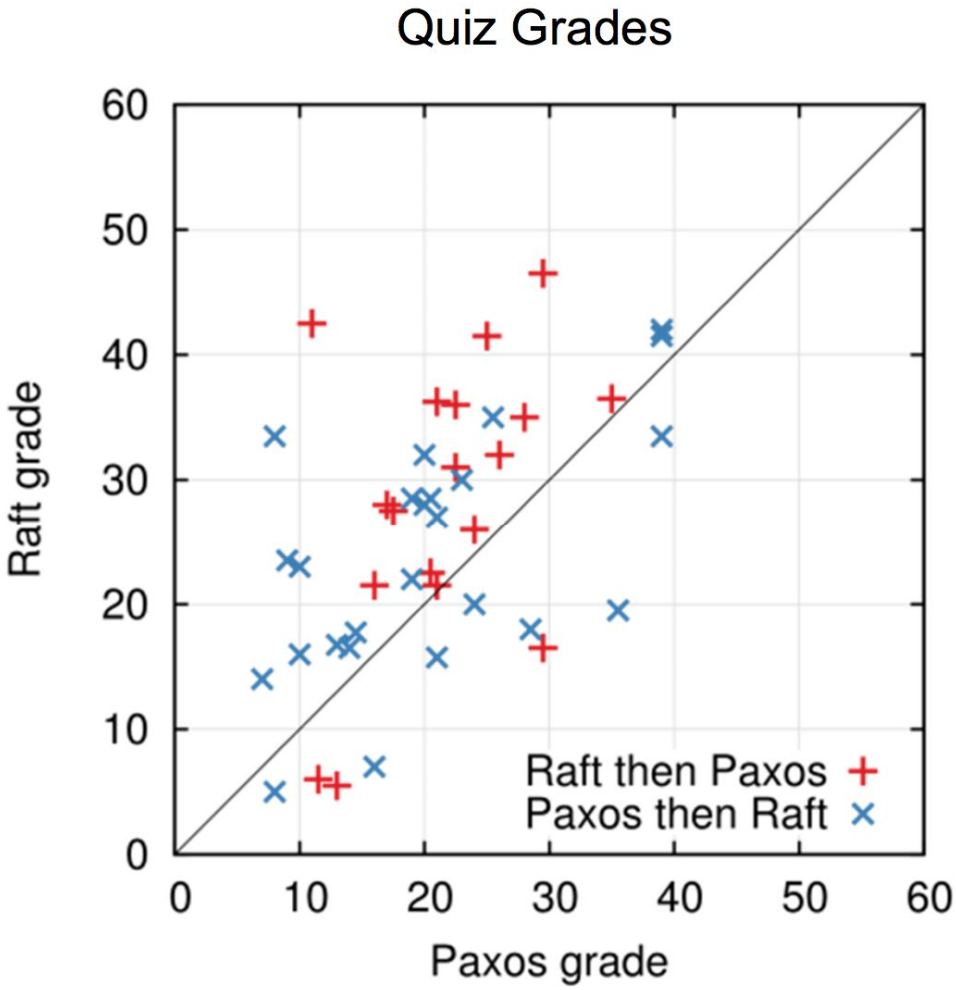
- Configuration Change
  - Cluster enters joint consensus state
- Log Compaction
- Client Connection
- Linearizability

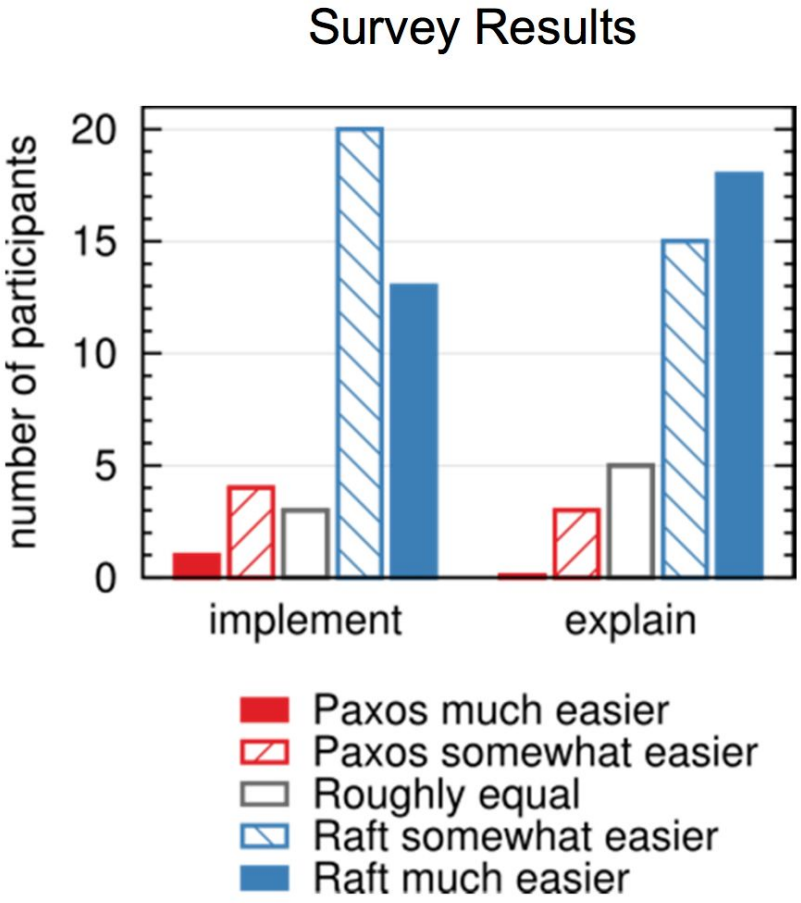


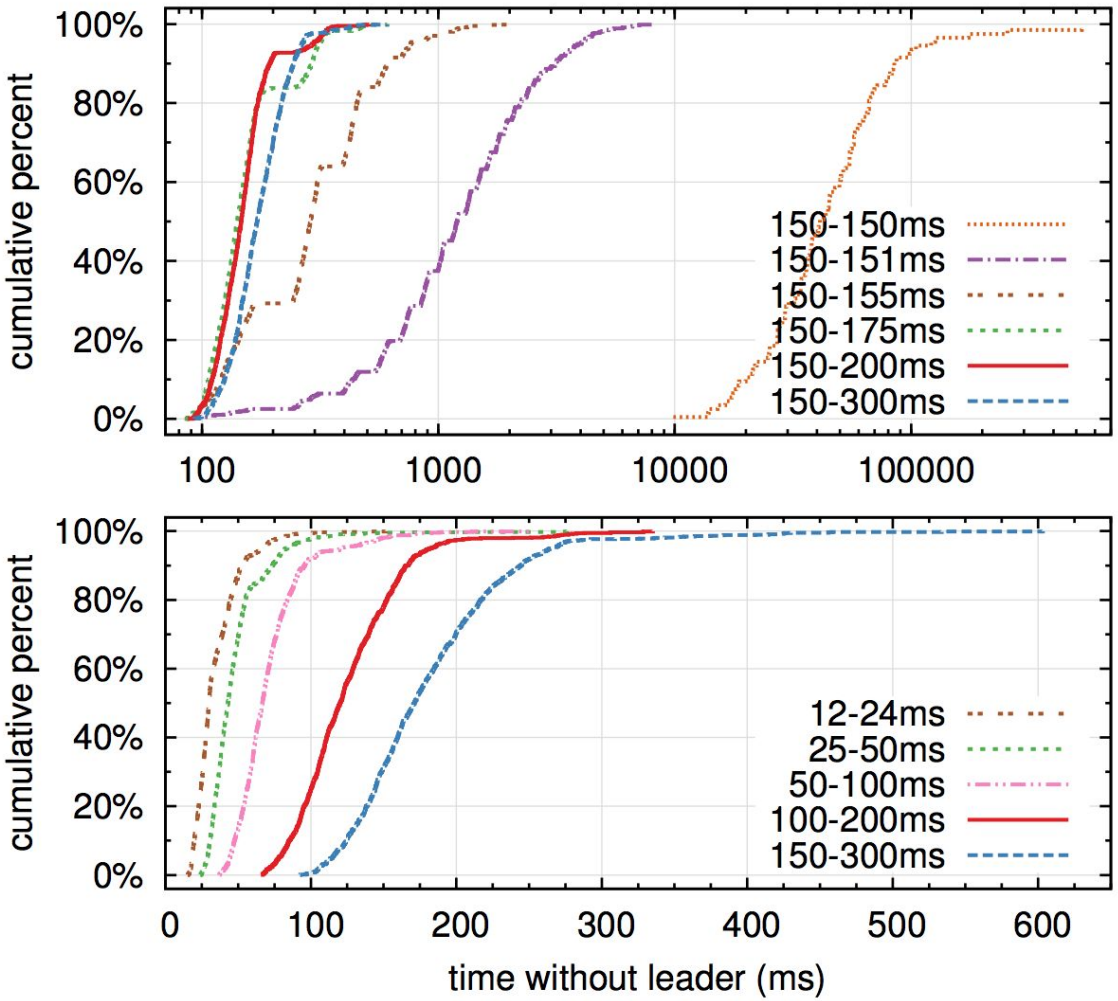
Demo

Were they successful?

Appendix


Quiz Grades
Raft grade
Paxos grade
0
10
20
30
40
50
60
Raft then Paxos
Paxos then Raft


Survey Results
number of participants
0
5
10
15
20
implement
explain
Paxos much easier
Paxos somewhat easier
Roughly equal
Raft somewhat easier
Raft much easier


cumulative percent
100%
80%
60%
40%
20%
0%
100
1000
10000
100000
150-150ms
150-151ms
150-155ms
150-175ms
150-200ms
150-300ms
cumulative percent
100%
80%
60%
40%
20%
0%
0
100
200
300
400
500
600
time without leader (ms)
12-24ms
25-50ms
50-100ms
100-200ms
150-300ms