How a typical application becomes distributed
How a typical application becomes distributed
SMR Reconfiguration

• Overview SMR
• (Vanilla) Paxos reconfiguration
• Contiguous log reconfiguration: VR, ZK, Raft
• Explicit reconfiguration: Cheap Paxos, Vertical Paxos, Virtually Synchronous Paxos, 700 BFTs
• Byzantine VP
State Machine Replication (SMR)
State Machine Replication (SMR)
State Machine Replication (SMR)

F-resilience
State Machine Replication (SMR)

F-resilience
Glossary

• replica(s)
  • leader/coordinator
• command sequence
• clients
• learn
A replicated system after a stormy night
A replicated system after a stormy night

with leader ranks
Paxos: Two-Phase SMR Protocol

- First phase: leader replacement, amortized
- Second phase: replication
Contiguous log

• most recent log wins
Vertical Paxos
reconfiguration == leader replacement

hmmm..
reconfiguration

- intra-sequence
reconfiguration

- intra-sequence
Virtually Synchronous Paxos

- inter-sequence
- separate steady state from reconfiguration
- two consensus decisions: state and next
Vertical Paxos

- steady state is F+1
- aux reconfiguration master
- Primary backup: special case, implicit wedge-and-snap
- Cheap Paxos: special case (F+1)-of-(2F+1), F standby’s

- why is it vertical?
Each of the phases of Paxos may use non-intersecting quorums. Only quorums from different phases are required to intersect. Majority quorums are not necessary as intersection is required only across phases.
Flexible Paxos: Use-Cases
BVP: Byzantine Vertical Paxos

• non-repudiation: interest in one correct replica
• wedge and snap: can “prove” decision
BVP in the synch model

- **3-message-delay**
  - steady state 2F+1
    - client-leader
    - leader-all (signed)
    - all-client (signed)
    - proceed with 2F+1 signed echoes

- closing state for reconfig
  - synchronously probe 2F+1

- liveness
  - (2F+1)-of-(3F+1)

}\non-repudiate

}\proof
BVP in the synch model

- **4-message-delay**
- steady state F+1
  - client-leader
  - leader-all (signed)
  - all-all (signed)
  - all-client (signed)
  - proceed with F+1 signed echoes

- closing state for reconfig
  - XFT: synchronously probe F+1 by F+1

- liveness
  - (F+1)-of-(2F+1)
BVP in the asynch model

• w/TPM
  • similar to sync
  • steady state F+1
  • proof by HW attestation
A dynamic fault model

• interplay between adversary and system

• begin handover / end handover
  • begin handover to speculative new: |new| - F correct
  • end handover from current: all of current can be faulty
Take-aways

• going beyond Paxos
• reconfiguration-based approach
overflow

ask me questions here; otherwise I’ll go on.
Flexible Paxos: Use-Cases
Flexible Paxos: Use-Cases

- 3+3
- 2+2+2