Paxos / Multi-Paxos

Admin

- Midterm

  + During Class

  + Open Book
    - Papers
    - Notes
    - ... 

  + Covers EVERYTHING UP TO THIS CLASS
    - Know what the async model is
    - Fairness
    - Linearizability, seq. cnt
    - Raft, Paxos

  + Questions we discussed in class or show up in notes are a good benchmark

- Final Project Proposal

  \[ \rightarrow \text{Campuswire Post} \]

Where we are

- Raft
- Building Block: Quorum Intersection!

- Replication
  - Commit entry by replicating to at least a quorum
  - Goal: Committed entries will never be lost

- Leader election
  - Leader's log is authoritative
  - Goal: Any node elected leader has a log containing all committed entries

  How
  - Check log completeness when voting for a leader
  - Require votes from a quorum

Multi-Paxos
Building Block: Quorum Intersection
- Replication
  \[\Rightarrow \text{commit entry by replicating to at least a quorum}\]
  \[\Rightarrow \text{Goal: Committed entries will never be lost}\]
  \[\Rightarrow \text{Do not depend on leader election to achieve this goal.}\]
  \[\Rightarrow \text{Leader's log is not special.}\]

Paxos/Synod: Agreement on one log entry (slot/index)

Proposer
- Propose commands for the slot

Acceptors
- Vote on proposals to decide commit (or not)

Learner
- Implement state machine
1. **Phase 1**. Figure out what value to use

2. **Phase 1a**

   - **Proposer**
   - **Acceptors**
   - **Prepare**
   - **Proposal ID**
Reminder: Only looking at 1 slot

- Acceptors send Promise to Proposer if

  has not previously seen Prepare with higher proposal ID.

- Included value if previously accepted a value

Phase 1b
Promise/Prepare Response
- Proposal ID
- Prev Value is any
- Proposal ID for prev accepted value.
* Compute Value for Phase 2

- Wait for Phase 1b (Promise/Prep Resp) responses from quorum of acceptors

\[ \langle \text{Promise}, \text{Proposal ID}, \text{Prev. Accepted Value}, \text{PropID when Accepted} \rangle \]

- If no phase 1(b) message contains a value: Propose any value

- If one or more phase 1(b) message contains a value:
  - Must use value accepted w/ highest prop ID

Claim: A committed value will always be used for phase 2.
**Why?**

**Phase 2: Replicate Chosen Value**

- **T=1**
  - **Proposer**
  - **Acceptors**

- **T=2**
  - **Proposer**
  - **Acceptors**

**Phase 2a**

\[
\langle \text{Accept, Proposal ID, Value} \rangle
\]

At acceptor:\n
If Proposal ID = Prop ID or last last promise

**Accept** Value

Tell learners: Phase 2b message

Note: Need to count # of accepts
(P2b m6%) to decide if committed.

Mapping this back

Prop ID = 2

Proposer

Accept

Learn

Learn

Learn

Learn

P1a

Prop ID = 2

Proposer

Accept

Learn

Learn

Learn

Learn

Prop ID = 2

Proposer

Accept

Learn

Learn

Learn

Learn

P, 1-2
Q: When is command committed?

Q: When can a learner apply a command?

Q: Requirements for Proposal ID:
Multi-Paxos: Extending To Many Slots

- What is missing: Leader needs to know what index to use for new commands

  Equivalently: Leader needs to know what indices are used

For any index \( i \)

I(a) \( \langle \text{PROPOSE, PROPOSAL ID} \rangle \)

I(b) \( \langle \text{PROMISE, PROPOSAL ID, PREV. ACCEPTED VALUE, PROP ID WHEN ACCEPTED} \rangle \)

- Tie proposal ID to leadership term.

ProposID = 3

![Diagram showing proposer, acceptor, learner nodes and their interactions]
Merging logs

- Apply rule from before

\[
\begin{array}{ccc}
0 & 1 & 2 \\
1 & 1 & 1 \\
2 & 1 & 2 \\
1 & 1 & \\
\end{array}
\]

Merged

Proposal ID-3

\{\text{Proposer}, \text{0-3, Merged}\}

\{\text{Accept}, v_0, v_1, v_2\}
Q. When is it safe to execute a command logged by a previous leader?

Raft:

Multi-Paxos:

Why all this flexibility?
- Propose commands for the slot

  - Disk Paxos: Use processors as proposers, disk as acceptors
  - Disk as learners

- Mencius/ePaxos: Scale in asymmetric setting

- Vertical Paxos: Reconfigure without stopping by changing acceptor sets
  [Similar to Rft reconfig]
Things not mentioned

- Failure detection: Knowing when a new leader should be elected

  \[ \text{PML of leader leases?} \]

- Leader election: Who becomes leader?