Paxos Made Simple

Lamport

Thomas Marshall
Motivation

- We need a way to maintain consistency in a distributed system in the presence of failures.
- 2PC works, but can get “stuck”, so a consensus algorithm is better.
Background

- Jim Gray proposes 2PC in the 1970s, but it blocks on single node failures.
- Dale Skeen proposes 3PC in the 1980s, but it produces incorrect results in some situations.
- Leslie Lamport proposes Paxos in 1998; the original paper describes the ancient Greek civilization on the Paxos island.
Example

- Leader election (eg. Mesos and Zookeeper) – important to only have one leader at a time.
- Some node(s) propose to be leader, other nodes can accept or reject.
Paxos

- We want to choose a value and have every node in the cluster agree on the value.
- Three classes of agents: proposers, acceptors, learners.
- Failures are possible, but non-Byzantine.
Safety Properties

- Only a value that has been proposed may be accepted.
- Only a single value is chosen.
- An agent never learns that a value has been chosen unless it actually has been.
Algorithm

- A proposer selects a proposal number $n$ and sends a request to the acceptors.
- If an acceptor has not already accepted a proposal with number greater than $n$, it responds that it can accept this proposal.
Algorithm (cont)

- If a proposer receives ready responses from a majority of acceptors, it sends an accept message.
- An acceptor that receives an accept message accepts the proposal unless it has responded to a prepare with a number greater than $n$. 
Paxos Example

proposer

acceptors

prepare = 8

prepare = 8

prepare = 8
Paxos Example

proposer

acceptors

highest proposed = 8

highest proposed = 8

highest proposed = 8
Paxos Example

proposer

accept = (8, foo)

acceptors

highest proposed = 8

highest proposed = 8

highest proposed = 8
Paxos Example

accepted = (8. foo)

proposer

accepted = (8. foo)

accepted = (8. foo)

acceptors

highest accepted = (8, foo)

highest accepted = (8, foo)

highest accepted = (8, foo)
Paxos Example

proposer

success = (8, foo)

success = (8, foo)

success = (8, foo)

leaners
Paxos Example (2)

acceptors

prepare = (8, A)

prepare = (8, A)

prepare = (8, A)

proposers

A

B
Paxos Example (2)

proposers

A
B

acceptors

highest proposed = (8, A)

highest proposed = (8, A)

highest proposed = (8, A)
Paxos Example (2)

acceptors

prepare = (9, B)

prepare = (9, B)

prepare = (9, B)

prepare = (9, B)

prepare = (9, B)

highest proposed = (8, A)

highest proposed = (8, A)

highest proposed = (8, A)
Paxos Example (2)

- Proposers: A, B
- Acceptors: highest proposed = (8, A), highest proposed = (9, B)
- A and B are ready to propose.

Diagram shows:
- Proposers A and B
- Acceptors with highest proposed values
- Arrows indicating readiness
Paxos Example (2)

Acceptors

Proposers

A

B

accept = ((8, A), foo)

accept = ((8, A), foo)

accept = ((8, A), foo)

highest proposed = (9, B)

highest proposed = (9, B)

highest proposed = (9, B)

highest proposed = (8, A)
Paxos Example (2)

proposers

accepted = ((8, A), foo)

A

B

acceptors

highest accepted = ((8, A), foo)

highest proposed = (9, B)

highest proposed = (9, B)
Paxos Example (2)

Proposers:
- A
- B

Acceptors:
- highest accepted = (8, A), foo
- highest proposed = (9, B)

Accept = ((9, B), bar)
Paxos Example (2)

proposers

A

B

acceptors

accepted = ((9, B), bar)

highest accepted = ((8, A), foo)

highest accepted = ((9, B), bar)

highest accepted = ((9, B), bar)

highest accepted = ((9, B), bar)
Paxos Example (2)

proposers

A

B

prepare = (10, A)

prepare = (10, A)

prepare = (10, A)

acceptors

highest accepted = ((8, A), foo)

highest accepted = ((9, B), bar)

highest accepted = ((9, B), bar)
Paxos Example (2)

proposers

A

B

acceptors

highest accepted = ((8, A), foo)
highest proposed = (10, A)

highest accepted = ((9, B), bar)
highest proposed = (10, A)

ready = ((10, A), ((8, A), foo))
ready = ((10, A), ((9, B), bar))
ready = ((10, A), ((9, B), bar))
Paxos Example (2)

A proposer makes a proposal (10, A) which is accepted by the acceptors:

- highest accepted = ((8, A), foo)
- highest proposed = (10, A)
- highest accepted = ((9, B), bar)
- highest proposed = (10, A)
- highest accepted = ((9, B), bar)
- highest proposed = (10, A)

Accepter B then accepts the proposal (10, A), finalizing the process.
Paxos Example (2)

- **Proposers**
  - A
  - B

- **Acceptors**
  - highest accepted = ((10, A), bar)

- **Accepted**
  - accepted = ((10, A), bar)
  - accepted = ((10, A), bar)
  - accepted = ((10, A), bar)
Progress

- You can imagine “dueling” proposers that continually propose higher and higher proposal numbers without any ever being accepted.
- Solution: distinguished proposer.