The Paxos Register
Harry C. Li, Allen Clement, Amtaan and Ayer, Lorenzo Alvisi
The University of Texas at Austin

Consensus & Paxos

• Validity
• Integrity
• Agreement
• Termination

Part-time Parliament [Lamport '98]

Safety

• Validity
• Integrity
• Agreement
• Termination

Liveness

Frangipani [Thekkath et al. '97]

Byzantine Paxos

Part-time Parliament

Frangipani

Byzantine Paxos

Part-time Parliament

Frangipani

Byzantine Paxos

Part-time Parliament

Frangipani

Byzantine Paxos

Part-time Parliament

Frangipani

Byzantine Paxos

Part-time Parliament

Frangipani

Byzantine Paxos
Consensus & Paxos

Safety
- Validity
- Integrity
- Agreement
- Termination

Liveness

Part-time Parliament
Frangipani
Symantine Paxos
Disk Paxos
Deconstructing Paxos
Reconstructing Paxos
Active Disk Paxos
Symantine Disk Paxos

[Lamport '98]
[Thekkath et al. '97]
[Castro & Liskov '99]
[Gafni & Lamport '00]
[Boichat et al. '01]
[Boichat et al. '01]
[Chockler & Malkhi '02]
[Abraham et al. '04]

Byzantine Paxos

Fast Byzantine Paxos

[Castro & Liskov '99]
[Chockler & Malkhi '02]
[Abraham et al. '04]
[Martin & Ailini '05]
[Lamport '05]
Consensus & Paxos

- Safety
  - Validity
  - Integrity
  - Agreement
  - Termination

Liveness

Part-time Parliament
Frangipani
Symmetric Paxos
Disk Paxos
Destructuring Paxos
Reconstructing Paxos
Active Disk Paxos
Agreement & Termination
Symmetric Disk Paxos
Part Symmetric Paxos
Full Paxos
Hybrid Quorums
Chubby
Syneva

[Lamport '98]
[Thekkath et al. '97]
[Castro & Liskov '99]
[Gafni & Lamport '00]
[Boichat et al. '01]
[Boichat et al. '01]
[Chockler & Malkhi '01]
[Zhao et al. '02]
[Abraham et al. '04]
[Martin & Alonso '08]
[Lamport '00]
[Coeling et al. '05]
[Serrure '05]
[Neis et al. '07]

What Makes Paxos Paxos?

- Guarantee safety
- Provide liveness during periods of synchrony
- Use leaders to coordinate quorums

Is that it?

Understanding Paxos

Abstract Paxos
  - Register & Leader
  - Ranked Register
  - Alpha of Indulgent Consensus

[Abstract Paxos]
[Lampson '01]
[Boichat et al. '01]
[Chockler & Malkhi '08]
[Guerraoui & Raynal '07]
[Coeling et al. '05]
[Serrure '05]
[Neis et al. '07]

Crash and Byzantine

Abstract Paxos
  - Register & Leader
  - Ranked Register
  - Alpha of Indulgent Consensus

[Abstract Paxos]
[Lampson '01]
[Boichat et al. '01]
[Chockler & Malkhi '08]
[Guerraoui & Raynal '07]
Understanding Paxos

Crash and Byzantine

- Abstract Paxos [Lampson '01]
- Register & Leader [Boichat et al. '01]
- Ranked Register [Chockler & Malkhi '06]
- Alpha of Indulgent Consensus [Guerraoui & Raynal '07]

No register-based description encompassing both crash and Byzantine variants

Paxos Register

- Captures crash and Byzantine Paxos variants
- Focuses on most difficult part of Paxos
- Implements a write-once register
Paxos Register
- Captures crash and Byzantine Paxos variants
- Focuses on most difficult part of Paxos
- Implements a write-once register
  read \[\rightarrow\] write

Outline
- Motivation
- Paxos Register specification
- Crash implementation
- Byzantine implementation

Consensus & Paxos Register
Consensus & Paxos Register

Details

Details

Details

Details

Details

Details
write starts as \textit{partial}.

\textit{Partial} write may become \textit{visible}.

\textit{Visible} write may become \textit{total}.
Almost Regular

Consistency Semantics
read returns last written value or any value being concurrently written

Consistency Semantics
read returns last written value or any value being concurrently written

- Paxos Register redefines ‘concurrent’
- Regular w.r.t. modified partial order
  1. Writes ordered by increasing timestamp
  2. Total write precedes higher timestamped read
  3. Read precedes higher or same timestamped write
Modified Partial Order

2. Total write precedes higher timestamped read

read token \( (v=1, \text{ts}=0) \)

write \( v=a, \text{token} = (1,0) \)

read token \( (v=a/1, \text{ts}=1) \)

write \( v=a/b, \text{token} = (a/1,1) \)

decide a/b

world time

Write-Once Register
Write-One Register

Theorem: All total writes write the same value.

• Paxos variants share high-level protocol
• Paxos variants use common abstractions
  • Timestamps, tokens
  • Partial, total, and visible writes

Outline

• Motivation
• Paxos Register specification
• Crash Implementation
• Byzantine Implementation

Crash Model

• Asynchronous system
• Proposers, acceptors, learners
• At least one proposer correct
• Majority of acceptors correct
visible when an acceptor accepts it

read token (v=!, ts=0)
write v=a, token = (!,0)

total when a learner receives acks
from a majority of acceptors

read token (v=!, ts=0)
write v=a, token = (!,0)

token (v=!, ts=0)
read token (v=!, ts=0)
write v=a, token = (!,0)

read token (v=!, ts=0)
write v=a, token = (!,0)
Outline

- Motivation
- Paxos Register specification
- Crash implementation
- Byzantine implementation

Byzantine Model

- Asynchronous system
- Digital signatures
- Proposers, acceptors, learners
- At least one proposer correct
- More than two thirds acceptors correct
write $v=a$, token = $(\bot, 0)$

visible when a quorum of acceptors echo

visible when a quorum of acceptors echo

total when a quorum of acceptors ack
visible when a quorum of acceptors echo total when a quorum of acceptors ack

write v=a, token = (!,0)

write v=a, token = (!,0)

token (v=!, ts=1)

read
write v=b, token = (!,1)

write v=a, token = (!,0)

write v=a, token = (!,0)

token (v=!, ts=1)

read
write v=b, token = (!,1)

write v=a, token = (!,0)

write v=a, token = (!,0)

token (v=!, ts=1)

read
write v=b, token = (!,1)
write \( v_a, \text{token} = (\pi, 0) \)

\[ \text{read } v_b, \text{token} = (\pi, 1) \]

write \( v_a, \text{token} = (\pi, 1) \)

\[ \text{read } v_b, \text{token} = (\pi, 1) \]

write \( v_a, \text{token} = (\pi, 0) \)

\[ \text{read } v_b, \text{token} = (\pi, 1) \]

write \( v_a, \text{token} = (\pi, 0) \)

\[ \text{read } v_b, \text{token} = (\pi, 1) \]

write \( v_a, \text{token} = (\pi, 1) \)

\[ \text{read } v_b, \text{token} = (\pi, 1) \]
Byzantine consensus algorithm diagram.
Take Away Points

Paxos Register

- Captures similarities in Paxos protocols
- Hides implementation details
- Important abstractions:
  - tokens, visible writes, total writes