Paxos@SIGACT2001

> ***Paxos Made Simple. Leslie Lamport, SIGACT News, 2001.***

Overview

It explains the Paxos protocol and the Multi-Paxos protocol.

Paxos (with Voting Set)

Protocol

- 每个 acceptor 维护它所接受的 ballot **集合**
- Phase1b: 在对 $prepare(n)$ 的回复 $ack(n, b, v)$ 中, $b < n$. ("*The proposal with the highest number **less than** n that it has accepted, if any.*")

Correctness Proof

只需要证明 Consistent 性质.

证明思路: Paxos protocol => P2^c => P2^b => P2^a => P2 => Consistent

- Paxos protocol => P2^c
  - 根据 "Make Promise 规则", Original Paxos 显然满足该性质。
  - ***难点: 如何证明优化版本满足该性质?***
- P2^c => P2^b
  - 使用强数学归纳法
  - Quorum 的相交性 + Make Promise 规则
- P2^b => P2^a
  - 先"issue"后"accept"
- P2^a => P2
  - 先"accept"后"chosen"

- P2 => Consistent
    - proposals 的全序性

Optimizations (with Largest Vote)

Protocol

- 每个 acceptor 只需要维护它所接受过的**最大**的 ballot。
- Phase1b: “*…, then it responds to the request with a promise … and with the highest-numbered proposal (if any) that it has accepted.**”
- Phase2a: "v is the value of the highest-numbered proposal among the responses, …"

Optimized Paxos vs. Original Paxos

关于 Phase 1b 消息 $ack(n, b, v)$

显然, original Paxos 满足: $n > b$。
但是, optimized Paxos 并不满足 $n > b$。(TLA+ 已验证)

- [ ] TLA+ 给出的反例

关于 Phase 2b 中的 $accept(n)$

*acceptor 接受的 ballot 的序号并不是严格递增的。(TLA+ 已验证)

- [ ] TLA+ 给出的反例

关于 $P2^c$

Optimized Paxos 不再满足 $P2^c$。(TLA+ 已验证)

- [ ] TLA+ 给出的反例

失败的场景:
2 个Proposers p_1, p_2， 3 个 Acceptors a_1, a_2, a_3:

- p_2 提议 $Prop(2)$, 与 a_1, a_2 完成了 Phase1a, Phase1b, Phase2a。p_2 发送消息 $Accept(2, v_2)$ 给 a_3, a_3 接受 *但是没有 make promise。*
- p_1 提议 $Prop(1)$, 与 a_1, a_3 完成了 Phase1a, Phase1b，分别收到 $(a_1, 2, -1, \perp)$ 与 $(a_3, 1, 2, v_2)$。按照 Phase2a，p_1 选择 $v = v_2$。***但是v_2 并不是< 1 中最大的ballot 对应的值。实际上, 没有< 1 的，所以$v = v_1$。***

尽管 $P2^c$ 不满足，但由于 p_2已经完成 Phase2a， p_1 的消息 $Accept(1, v_1)$ 一定不会被 decided，所以不破坏 Consistency 性质。

这种场景下，p_1 在收到 $(a_3, 1, 2, v_2)$时, 可以 abort。更一般地, 如果收到的 Phase1b 消息 a_i, m, b, v_b 中, $b > m$, 该 proposer 可以 abort。

对 Optimized Paxos 的修改 (Research)

参考 PaxosStore@VLDB2016 与 DiskPaxos@DC2003, 将 Phase1b 与 Phase2b 合并:
有什么好处???

Phase 1b

允许 accept ballot

Phase 2b

允许 make promise

Multi-Paxos

> Written with StackEdit.