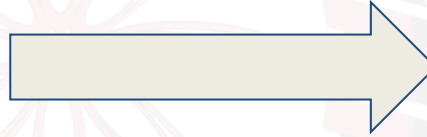


Why Distributed Consensus



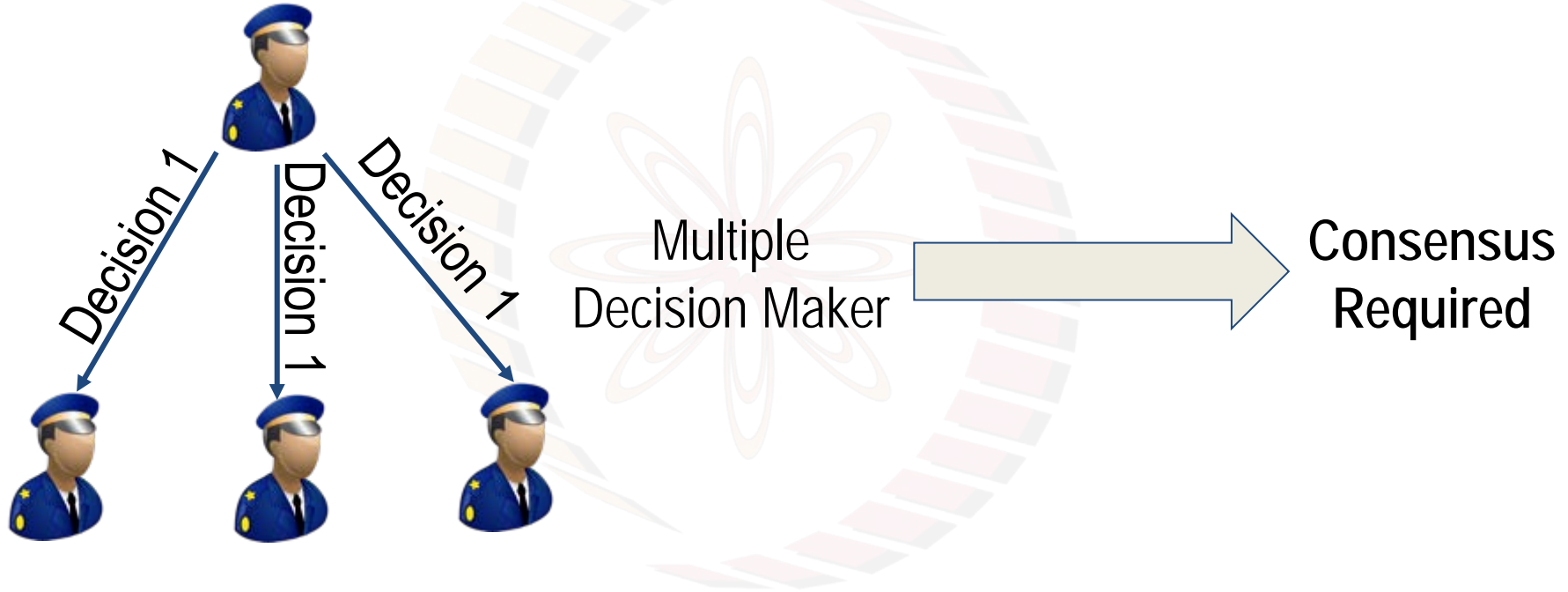
One Decision
Maker



No
Consensus



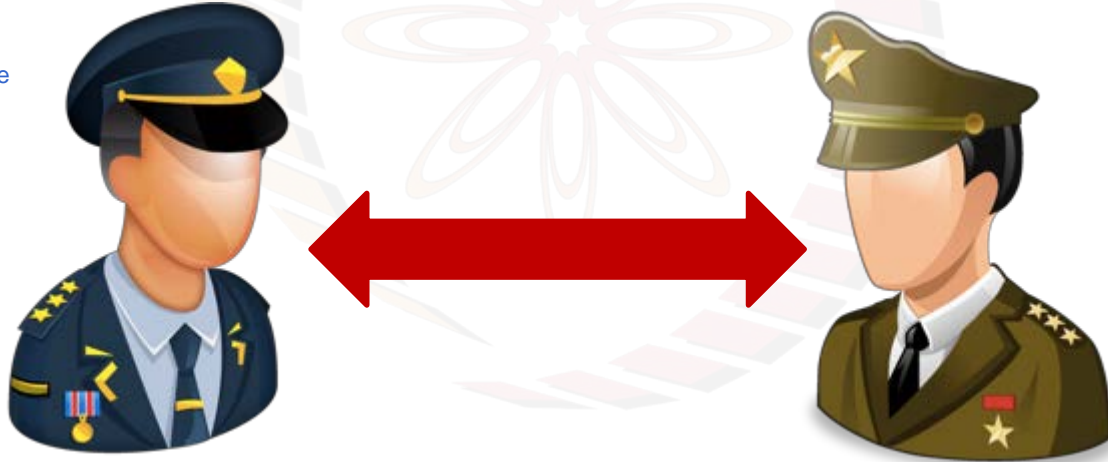
Why Distributed Consensus



Why Distributed Consensus

- So, no need of consensus in a single node process.
- **What about when there are two nodes?**
 - Network or partitioned fault, consensus cannot be reached

So we always need more than two nodes

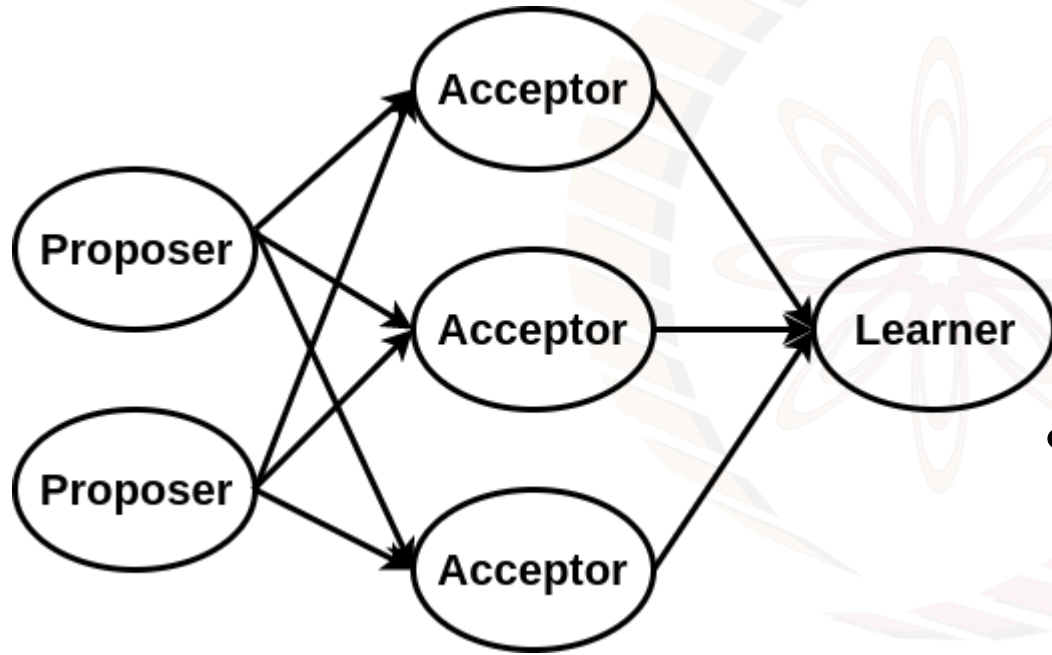


Different Algorithms

- **Crash or Network Faults:**
 - PAXOS
 - RAFT
- **Byzantine Faults (including Crash or Network Failures):**
 - Byzantine fault tolerance (BFT)
 - Practical Byzantine Fault Tolerance (PBFT)



PAXOS

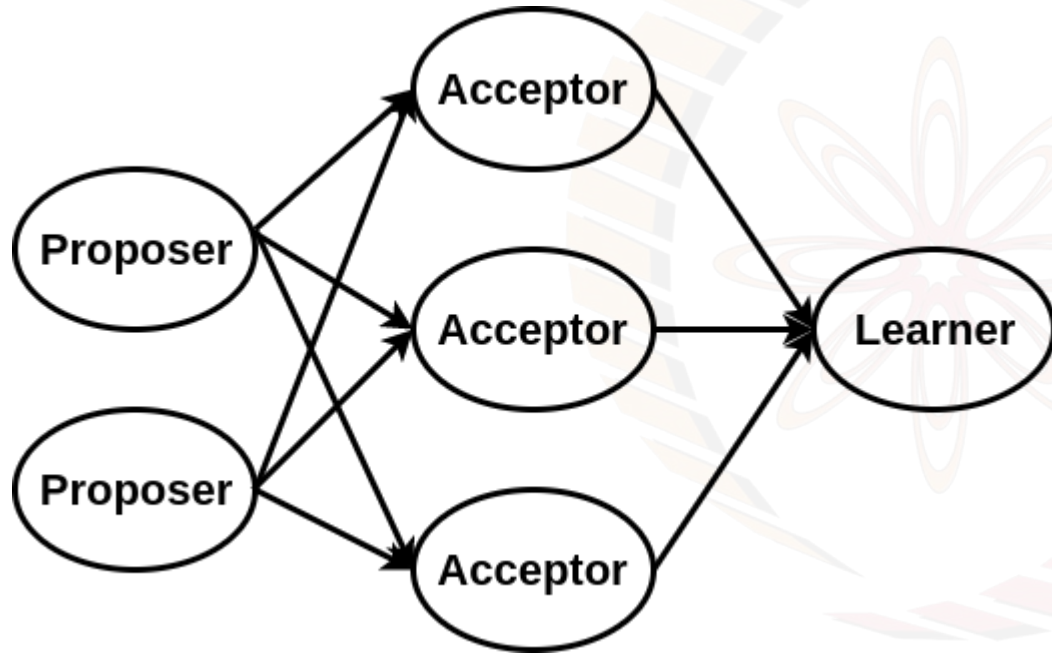


- System process
 - Making a proposal
 - Accepting a value
 - Handling Failures

Source: Lamport, Leslie. "Paxos made simple." ACM Sigact News 32.4 (2001): 18-25.

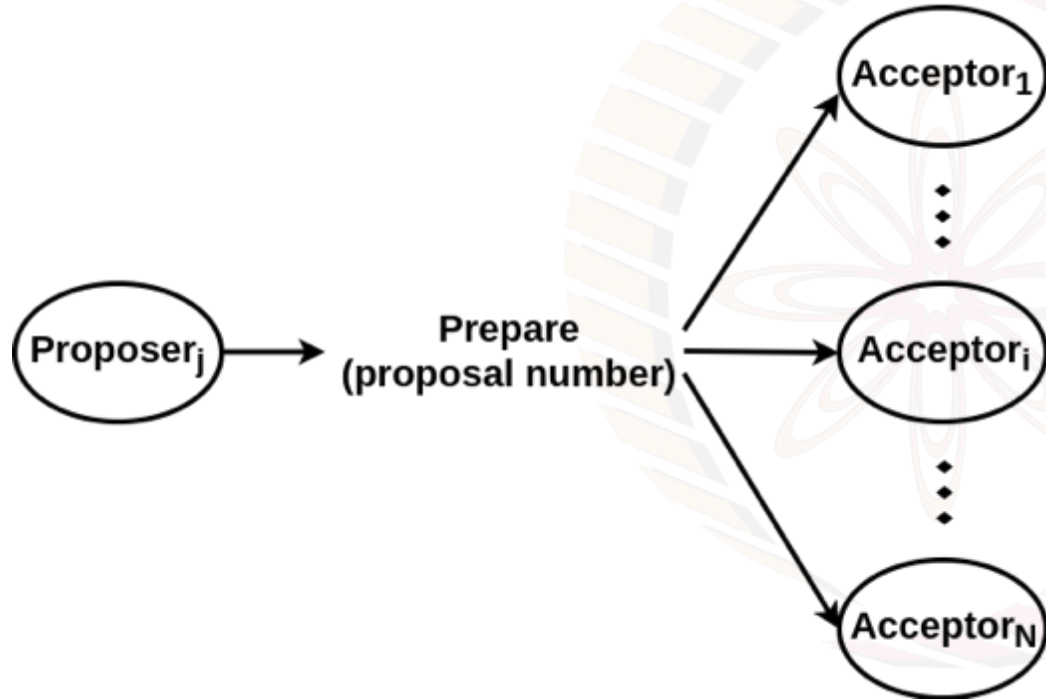


PAXOS: Types of Nodes



- **Proposer:** propose values that should be chosen by the consensus
- **Acceptor:** form the consensus and accept values
- **Learner:** learn which value was chosen by each acceptor

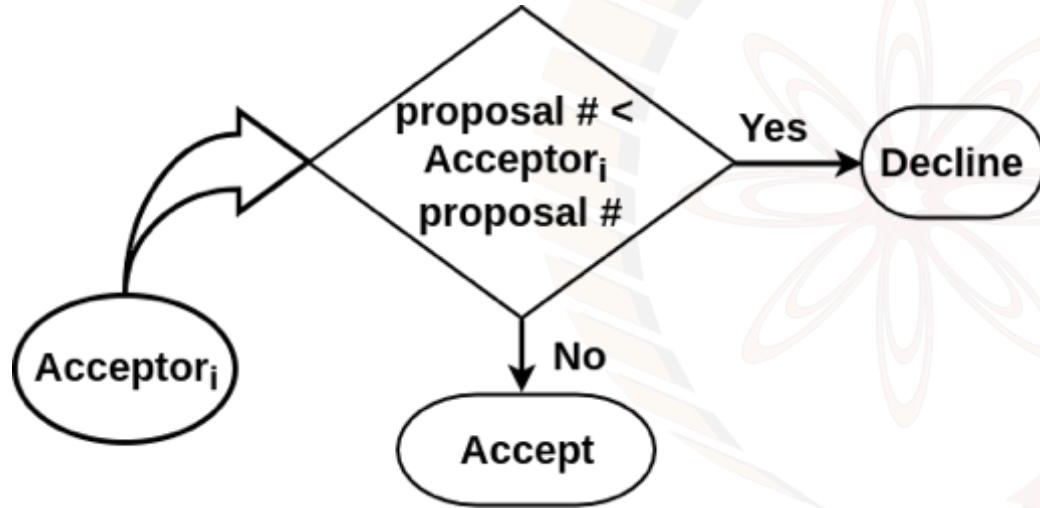
Making a Proposal: Proposer Process



- **proposal number:** form a timeline, biggest number considered up-to-date

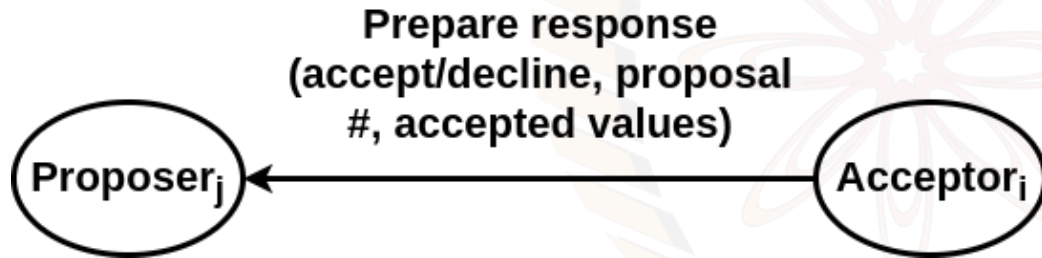


Making a Proposal: Acceptor's Decision Making



- Each acceptor compares received proposal number with the current known values for all proposer's prepare message

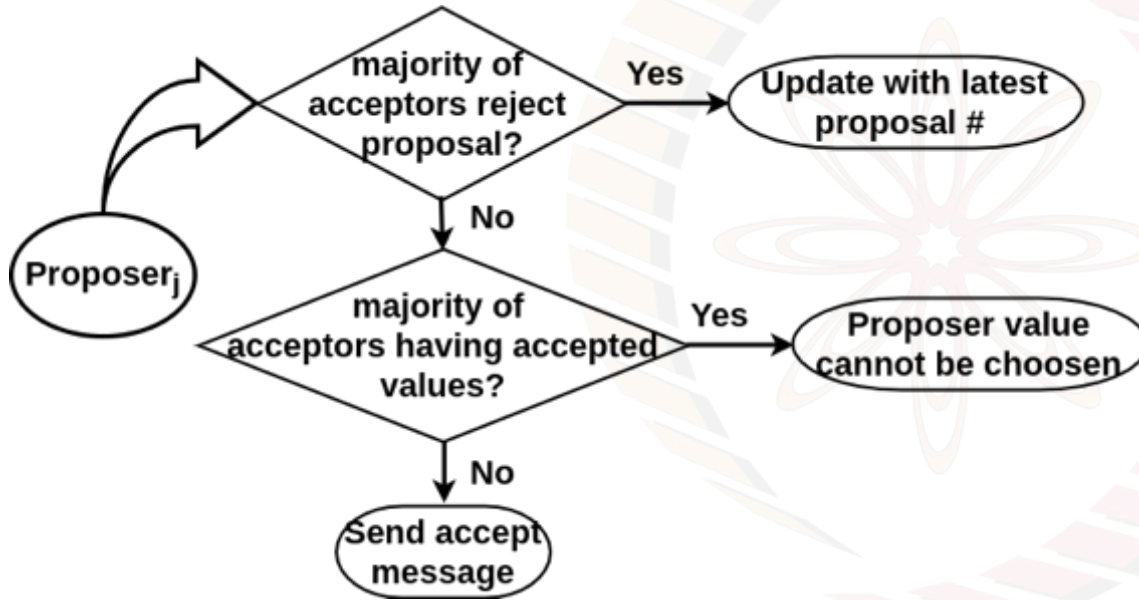
Making a Proposal: Acceptor's Message



- **accept/decline**: whether prepare accepted or not
- **proposal number**: biggest number the acceptor has seen
- **accepted values**: already accepted values from other proposer

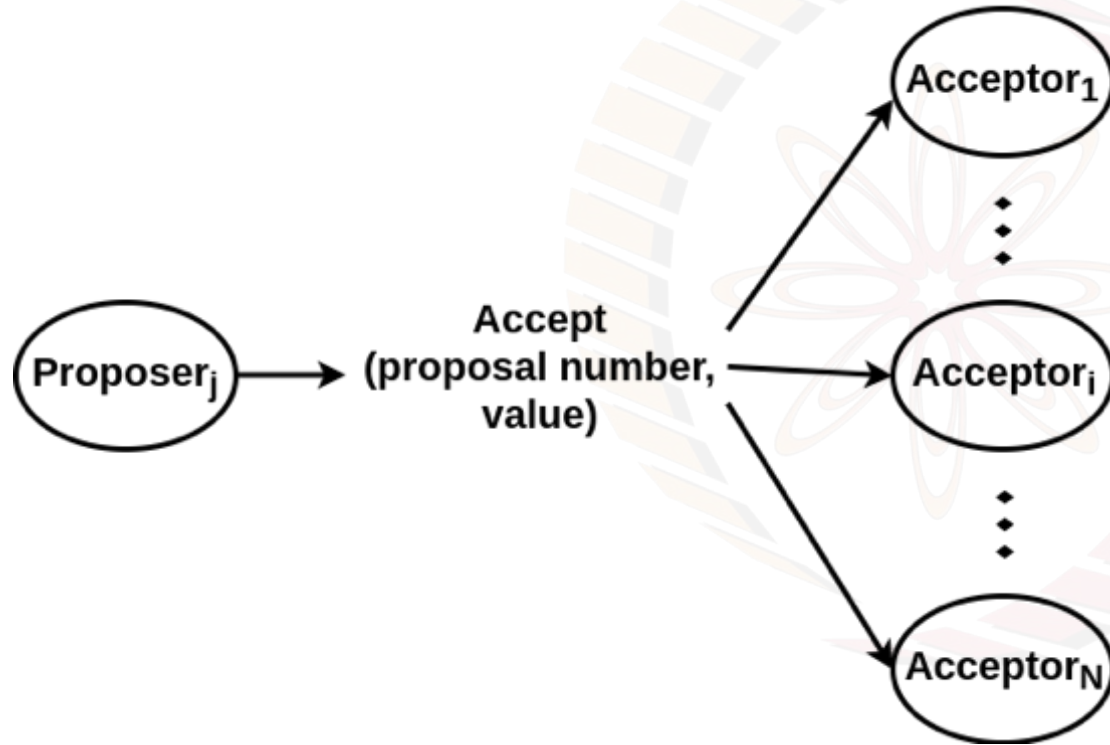


Accepting a Value: Proposer's Decision Making



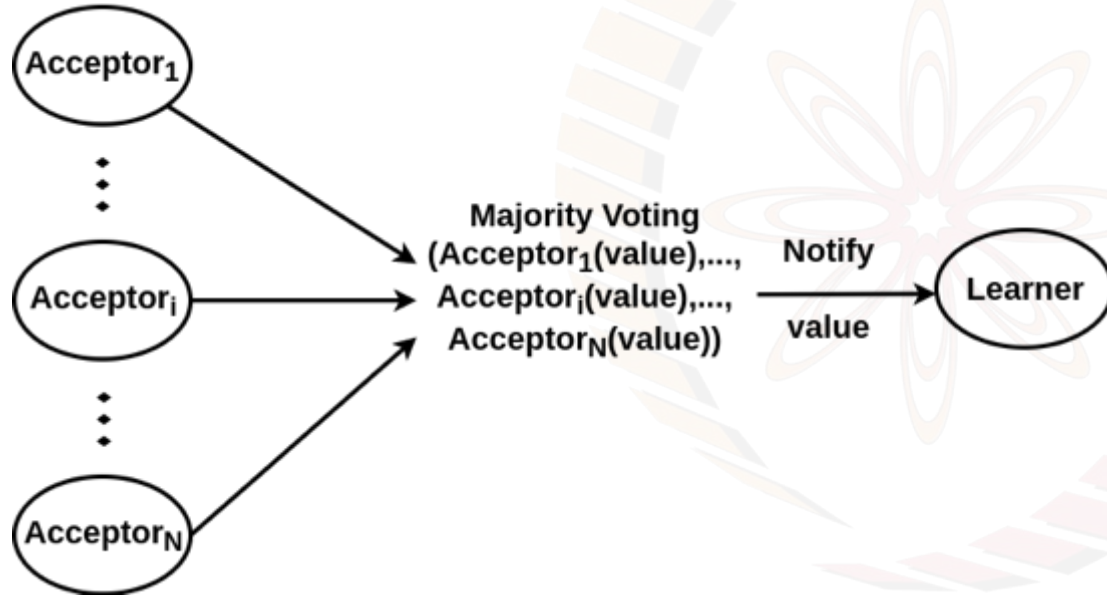
- Proposer receive a response from **majority** of acceptors before proceeding

Accepting a Value: Accept Message



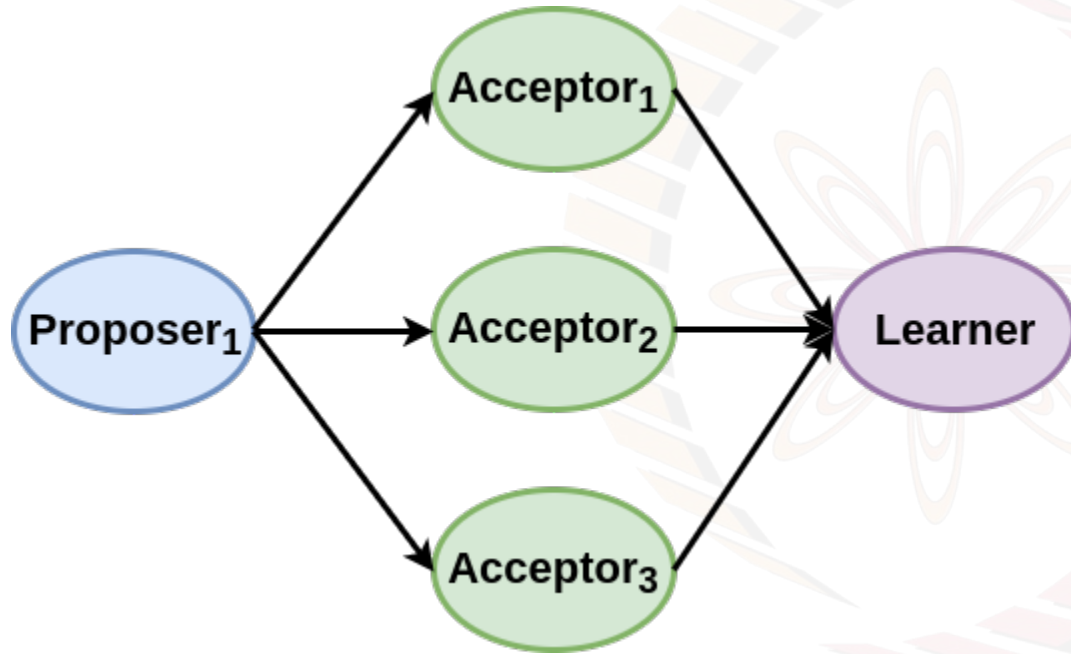
- **proposal number:** same as prepare phase value
- **value:** single value proposed by proposer

Accepting a Value: Notifying Learner



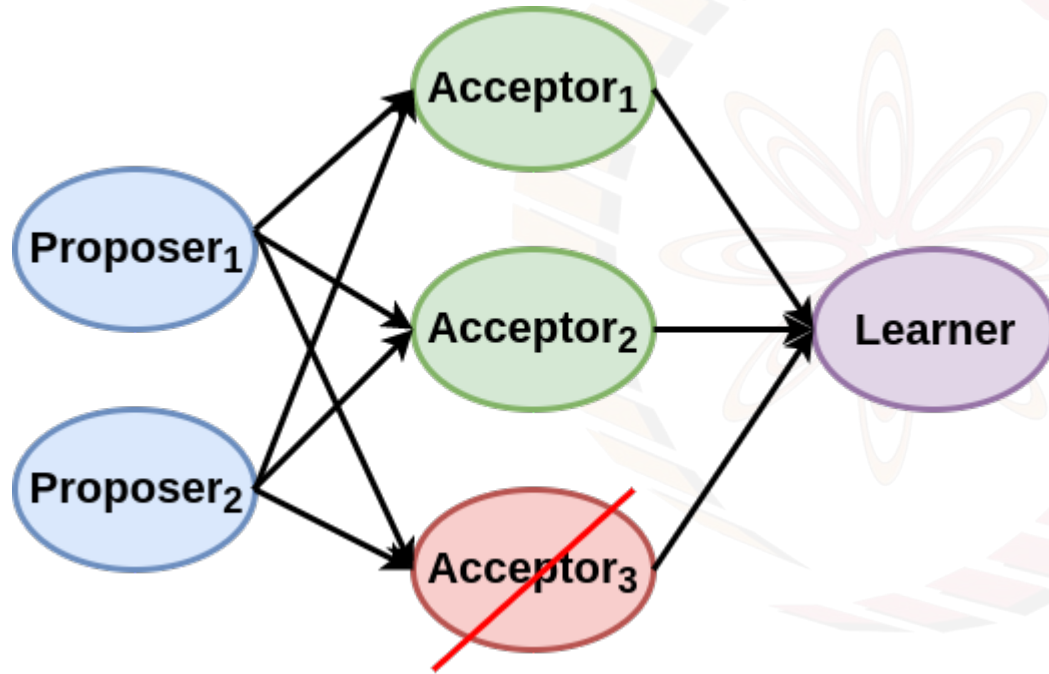
- Each acceptor accept value from any of the proposer
- Notify learner the majority voted value

Single Proposer: No Rejection



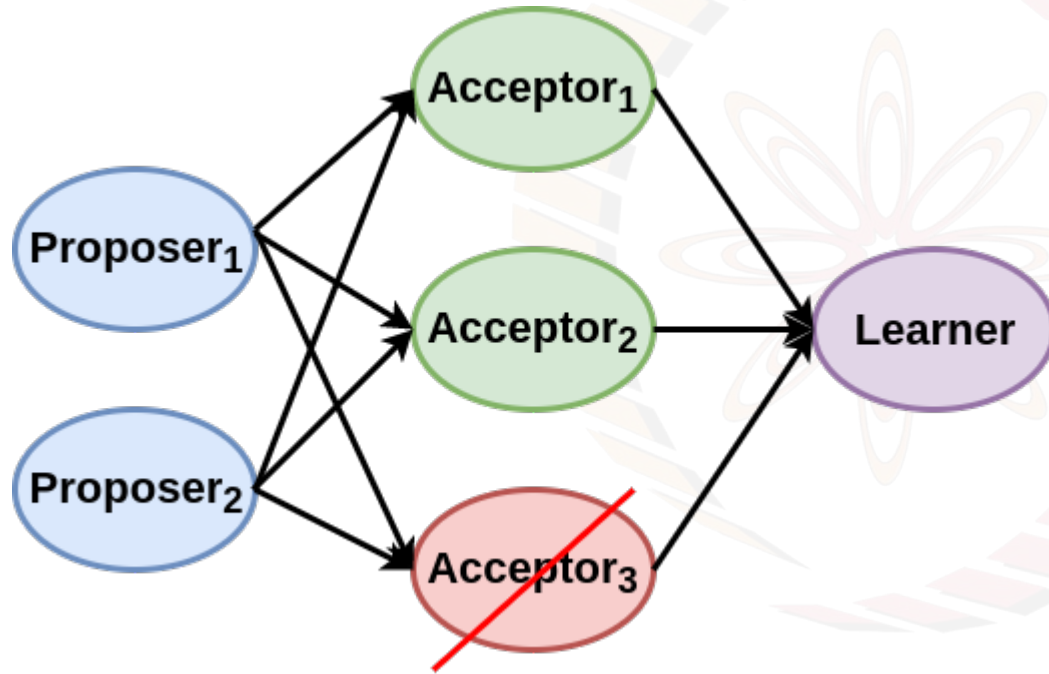
- Proposer always have proposal with biggest number
- No proposal rejected

Handling Failure: Acceptor Failure



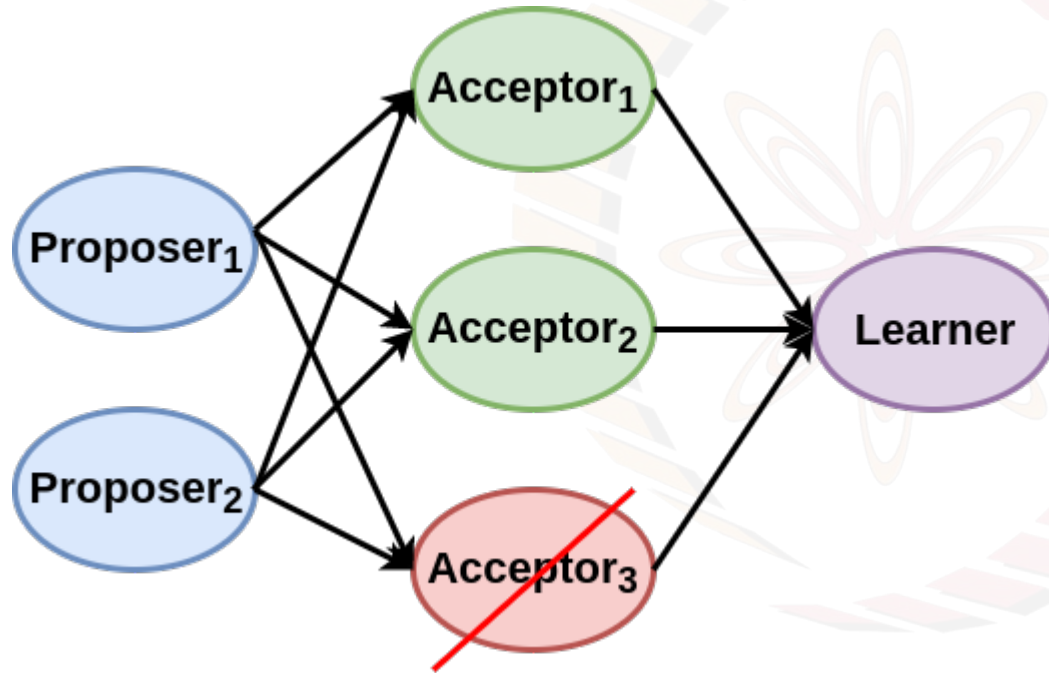
- Acceptor fails during prepare
 - No issues, other acceptor can hear the proposal and vote

Handling Failure: Acceptor Failure



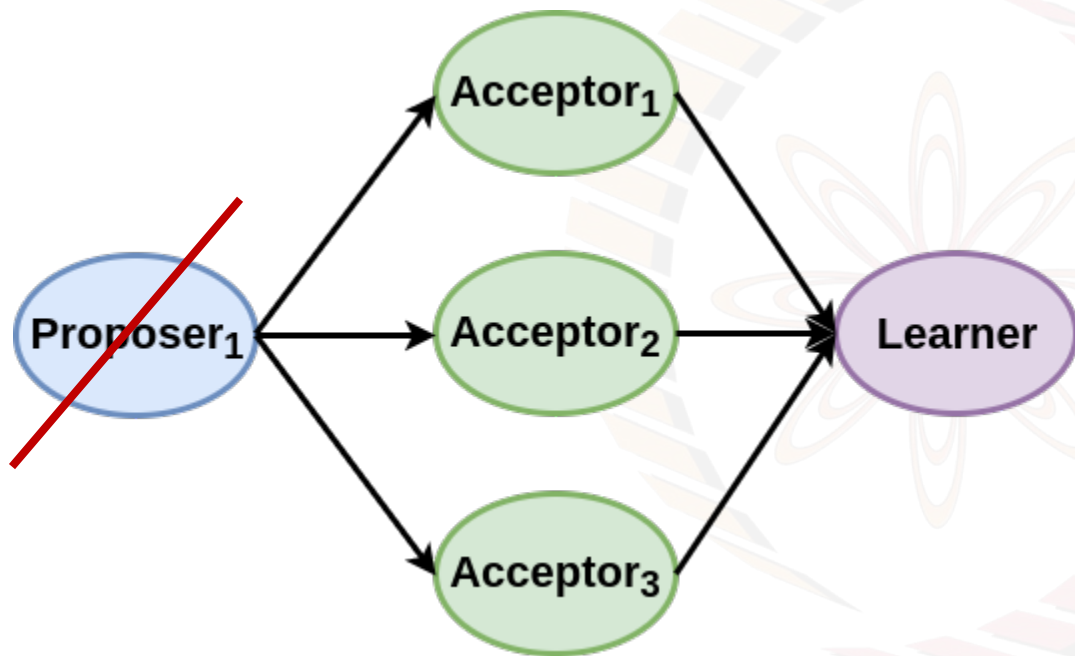
- Acceptor fails during accept
 - Again, no issues, other acceptor can vote for the proposal

Handling Failure: Acceptor Failure



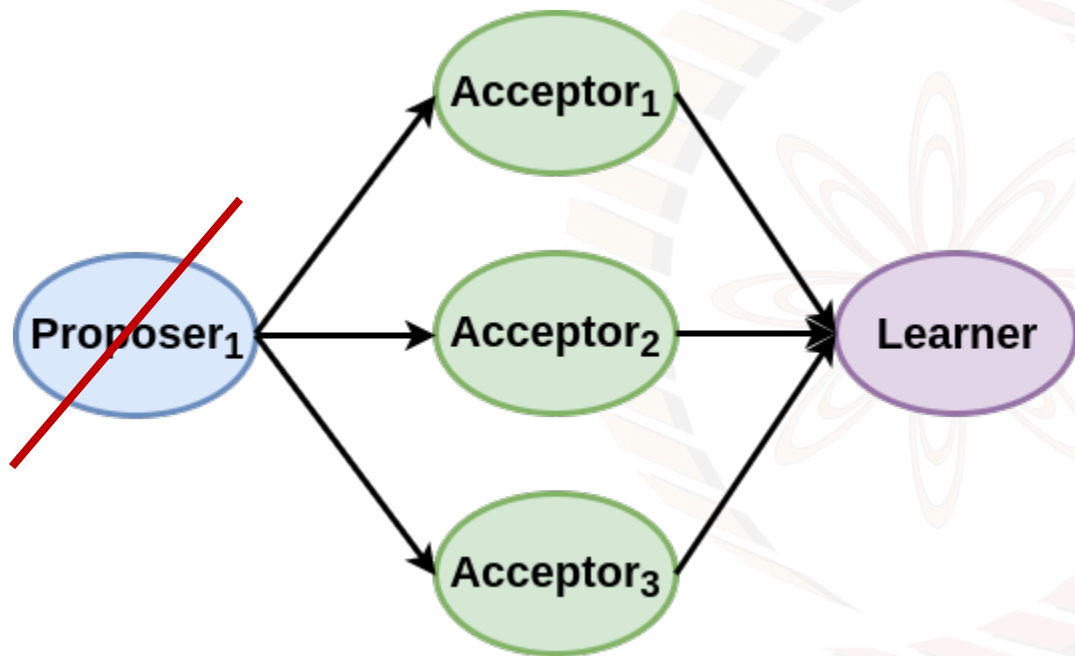
- More than $N/2 - 1$ acceptors fail
 - no proposer get a reply
 - no values can be accepted

Handling Failure: Proposer Failure



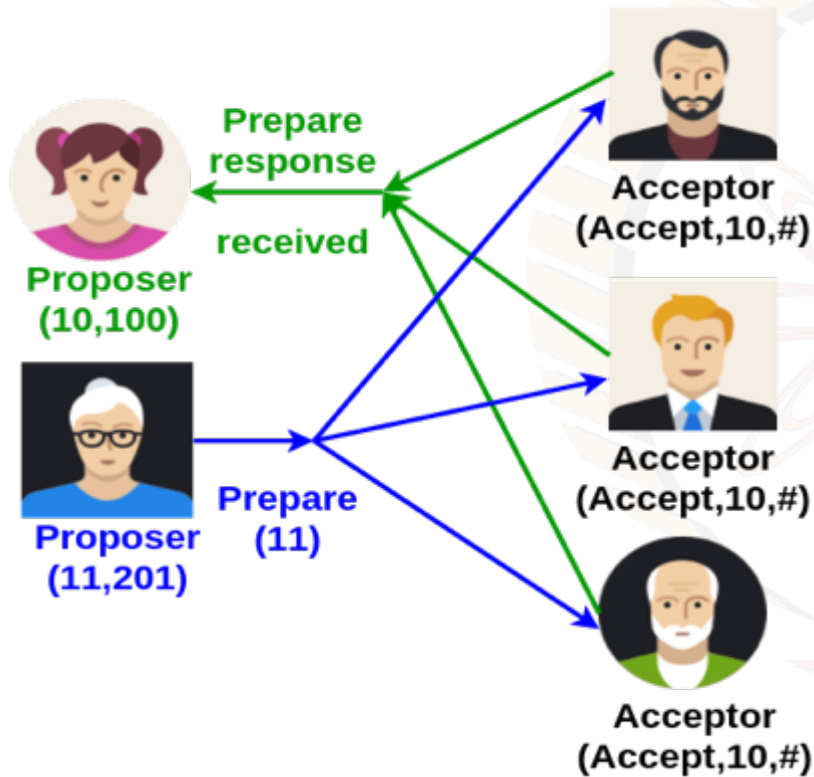
- Proposer fails during prepare phase
 - Acceptors wait, wait, wait, and then someone else become the proposer

Handling Failure: Proposer Failure



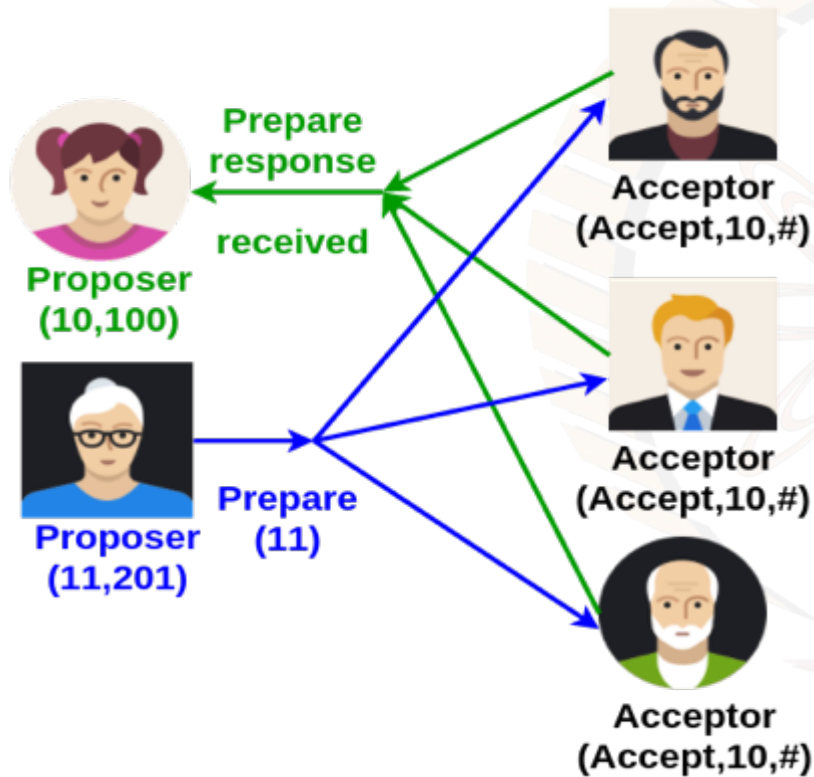
- Proposer fails during accept phase
 - Acceptors have already agreed upon whether to choose or not to choose the proposal

Handling Failure: Dueling Proposers



- Proposer received confirmations to her prepare message from majority
 - yet to send accept messages
- Another proposer sends prepare message with higher proposal number
- Block the first proposer's proposal from being accepted

Handling Failure: Dueling Proposers



- Use **leader election** - select one of the proposer as leader
- Paxos can be used for leader election !!

