Paxos Made Simple

Distributed Consensus

Rodgers Andati

4th Oct 2022

Background

Lynch & Liskov

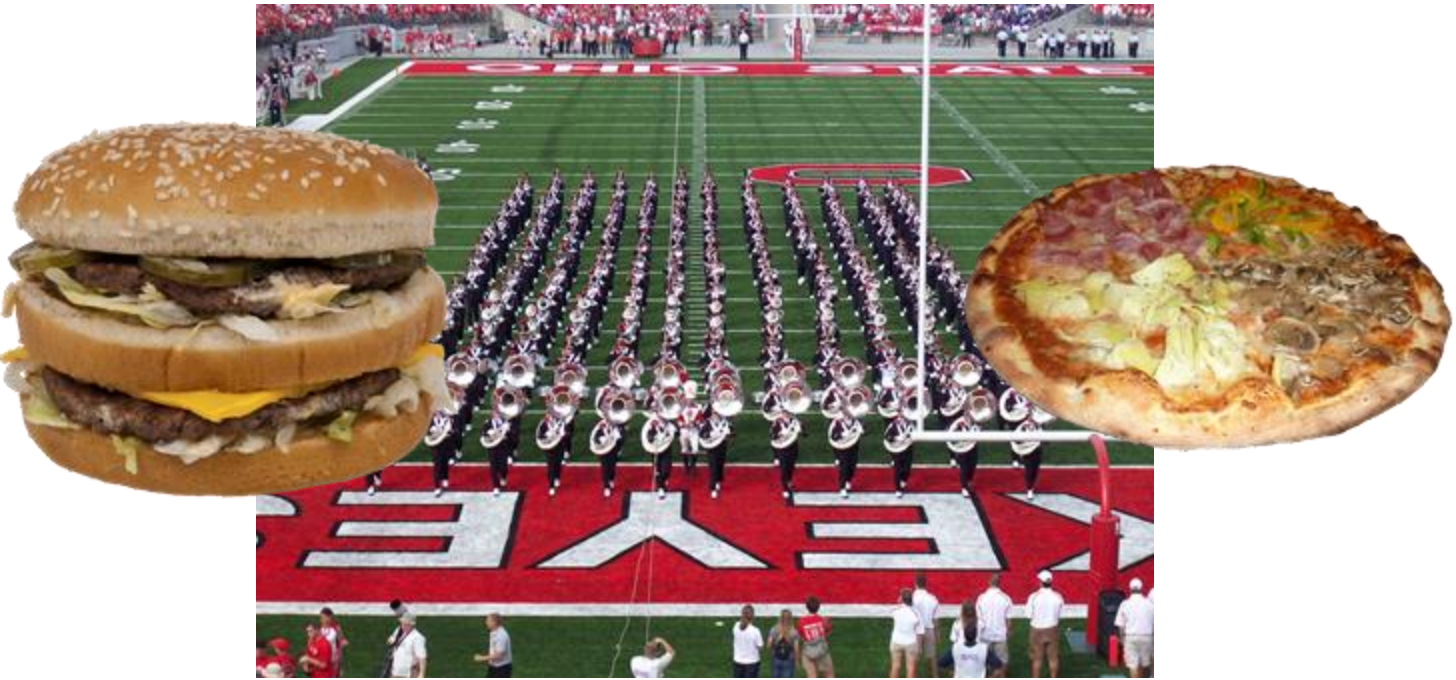
Leslie Lamport

The Part Time Parliament

Paxos Made Simple

Multi-paxos: Paxos + Complexity

Burgers or Pizza?



 or  ?

- Band director went home
- People easily distracted
- No fun if group splits up
- Hungry: must come to a decision fast!
- Yelling fails. Use person-to-person communication.

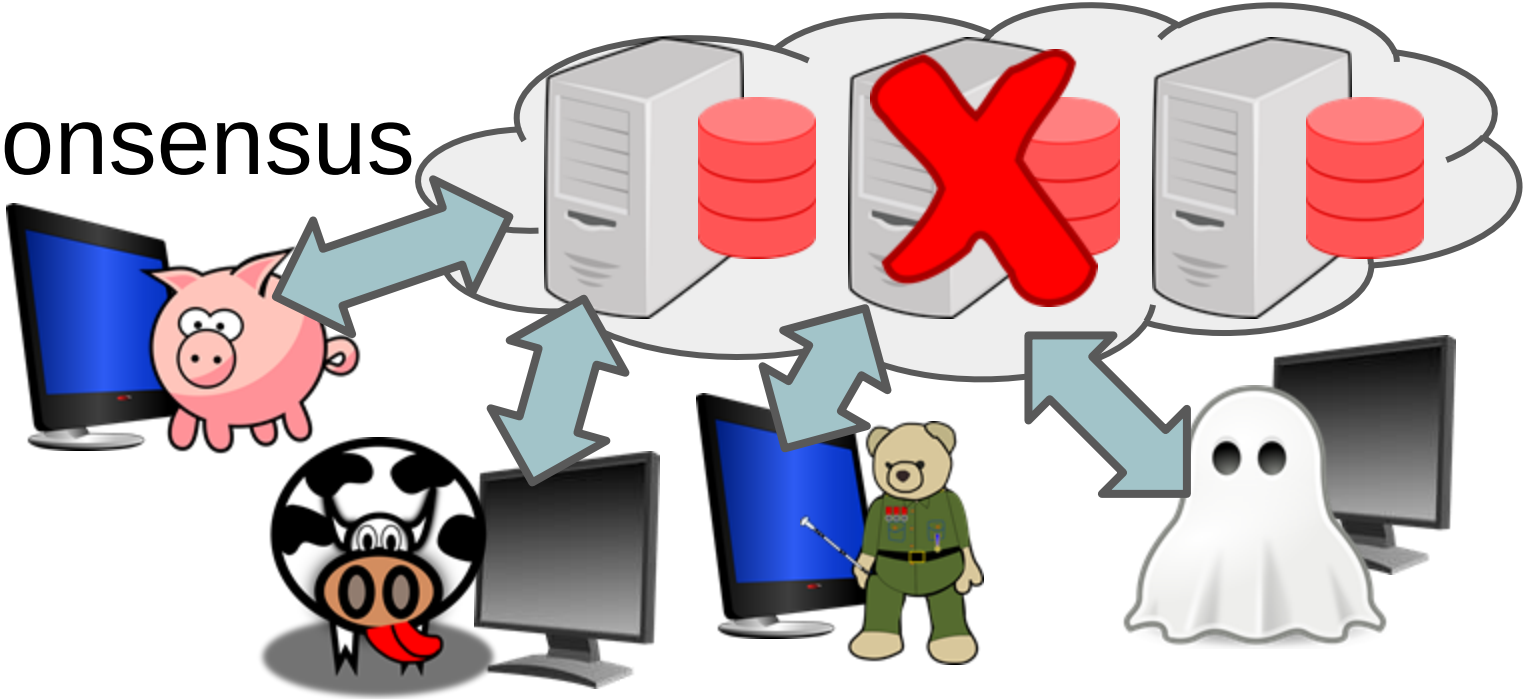
Want to achieve *consensus*

Burgers or Pizza?

1. What's happening?
2. Let's go for burgers!

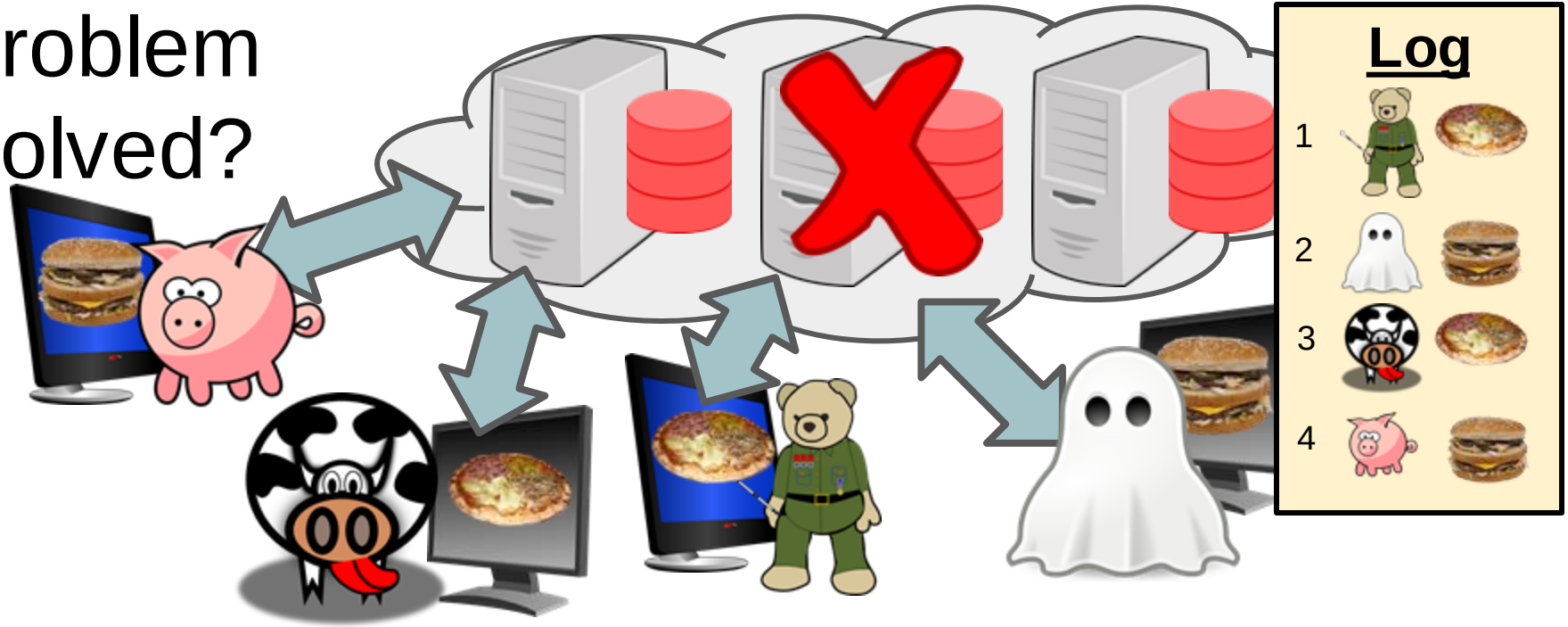
Paxos: *almost* this simple

Consensus



- ***If*** computers always agree
  - All computers are equivalent
  - Failure is no problem!
  - Living the dream
    - At the cost of some complexity…

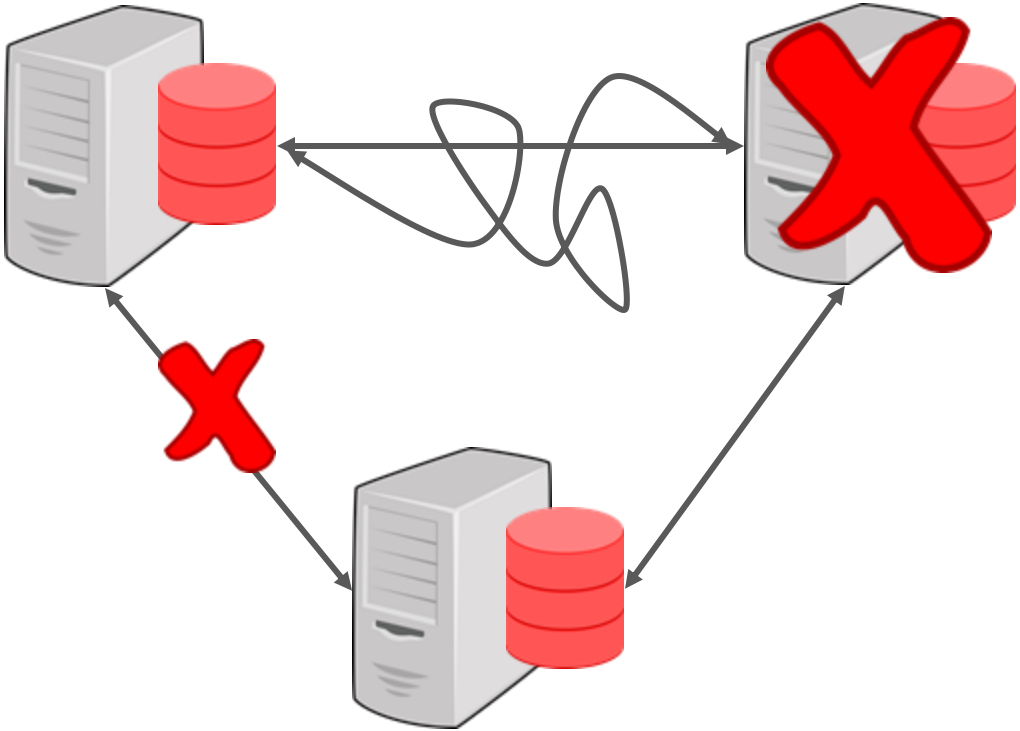
Problem Solved?



- Paxos lets you make *one choice*
- Multi-Paxos needed for a real system
- Build a *log* of choices

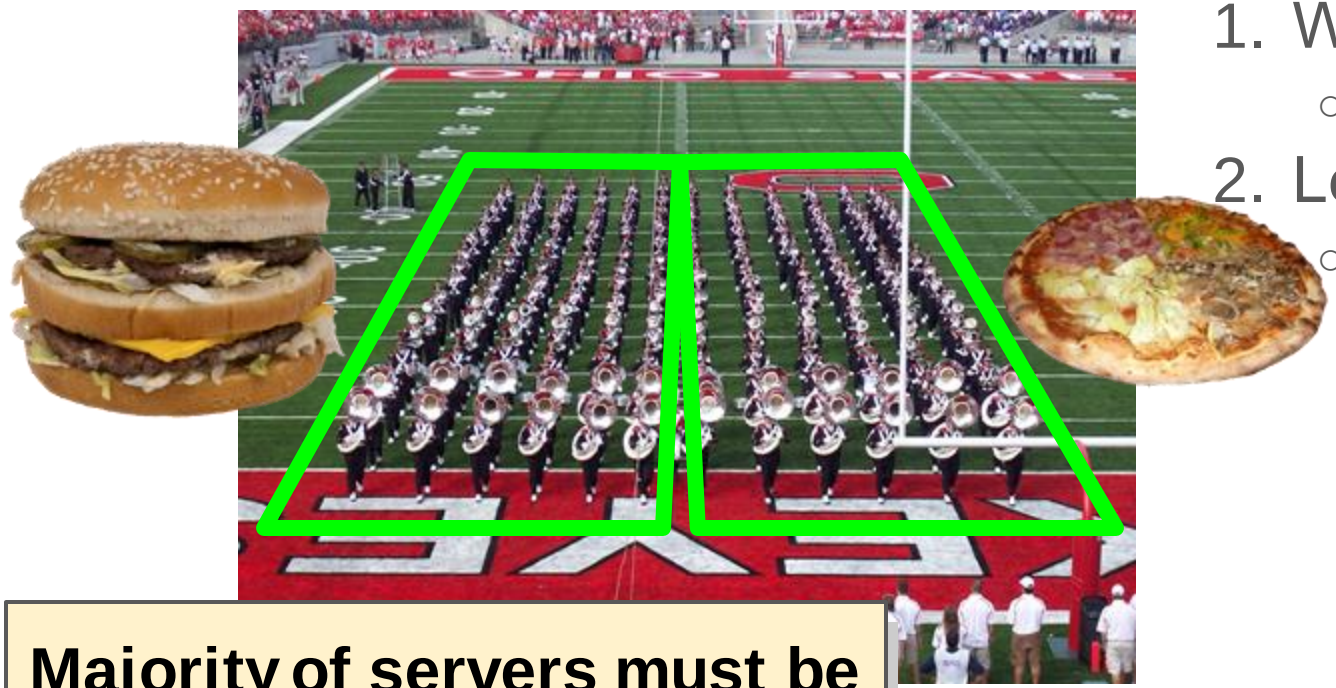
How does Paxos work?

Failure Model



Fail stop,
NOT Byzantine

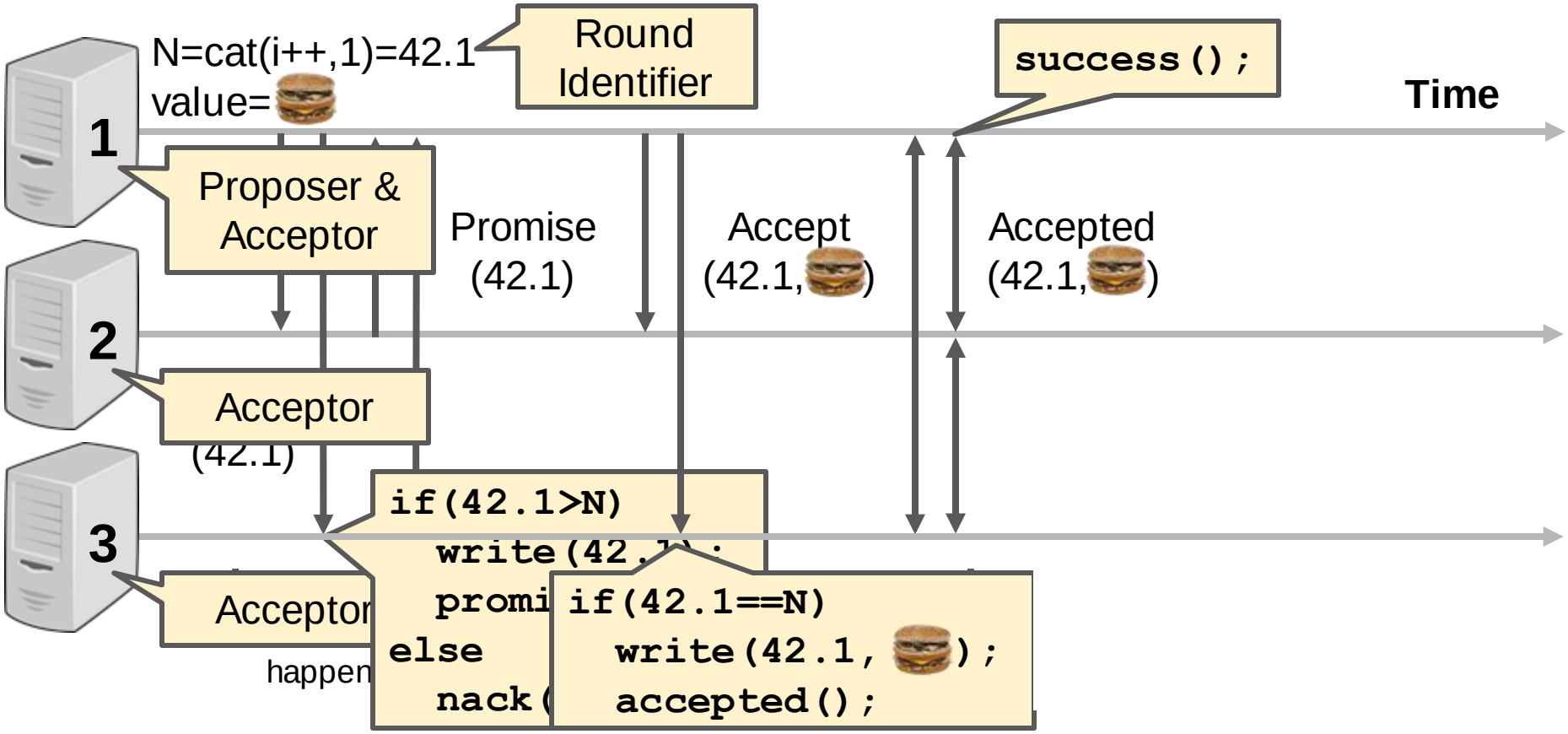
Majority wins!



1. What's happening?
   - Need to ask majority
2. Let's go for burgers!
   - Majority must agree

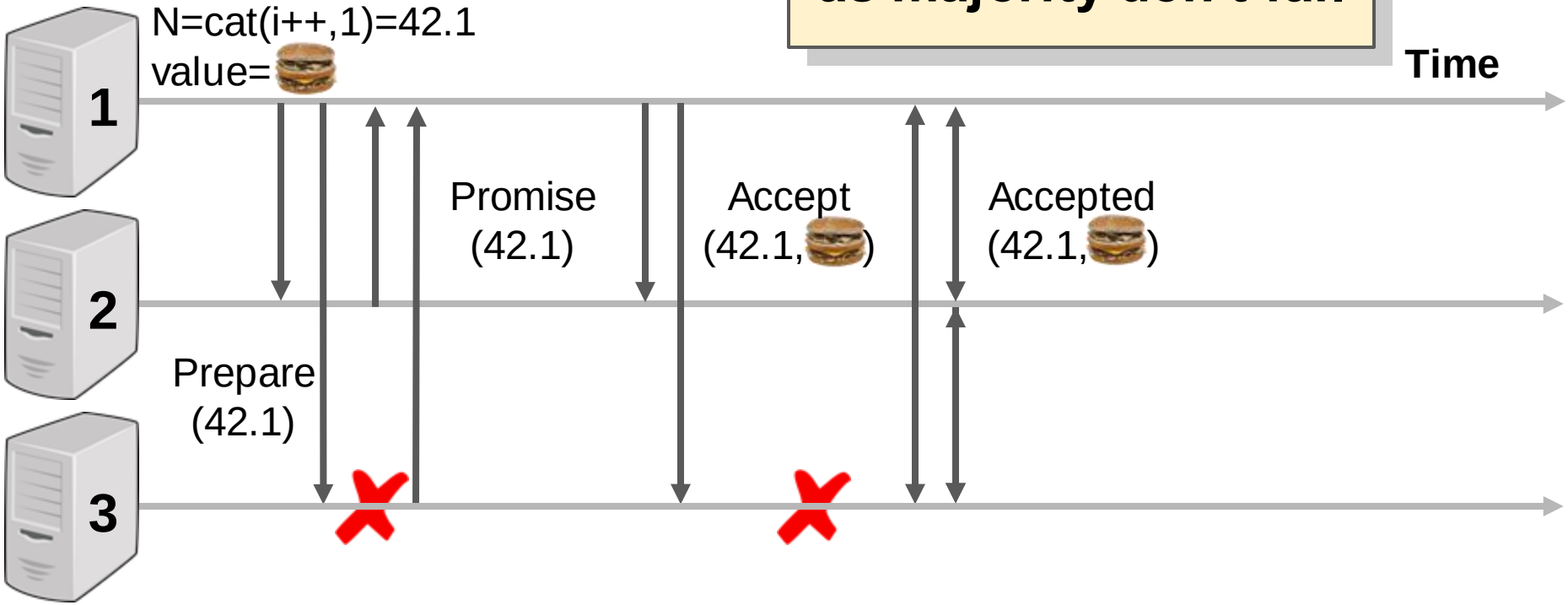
Majority of servers must be up for Paxos to terminate.

Basic Paxos

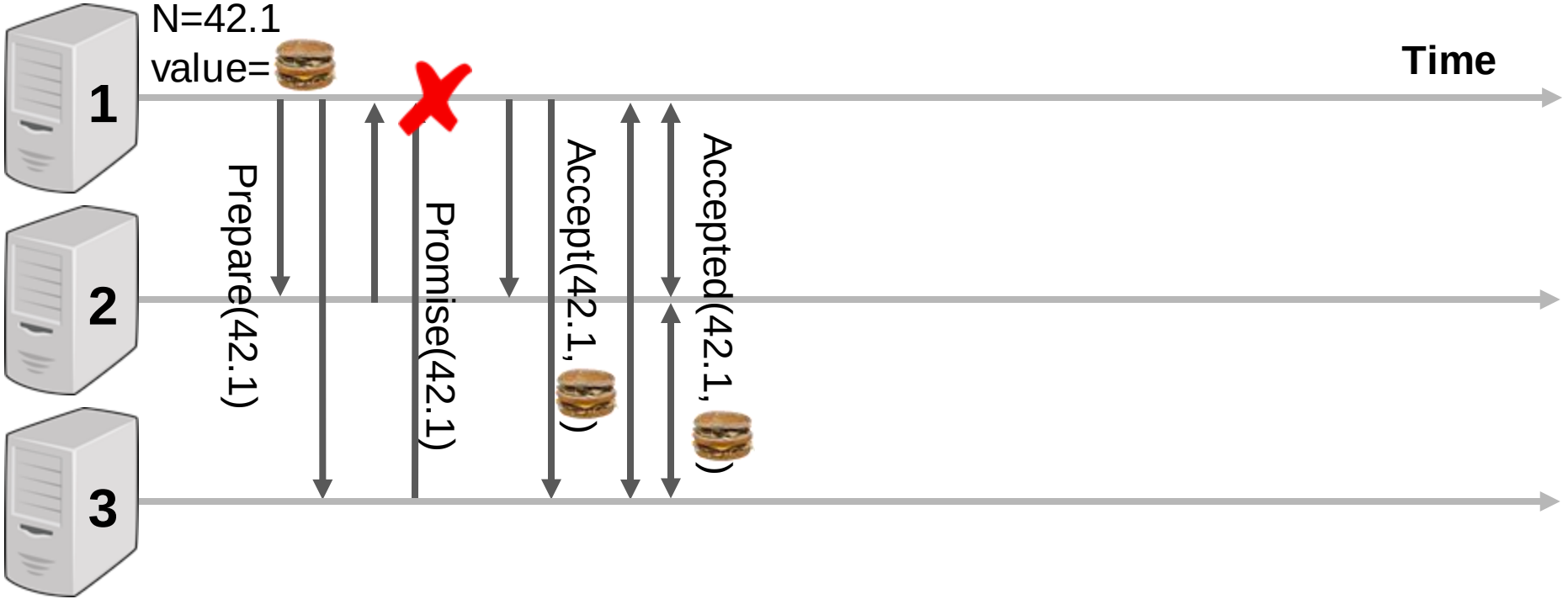

N=cat(i++,1)=42.1
value=
Round Identifier
success();
Time
1
Proposer & Acceptor
Promise (42.1)
Accept (42.1,
Accepted (42.1,
2
Acceptor
(42.1)
3
Acceptor
happen
if(42.1>N)
write(42.1);
promi
else
nack(
if(42.1==N)
write(42.1,
);
accepted();

Failures: Acceptor

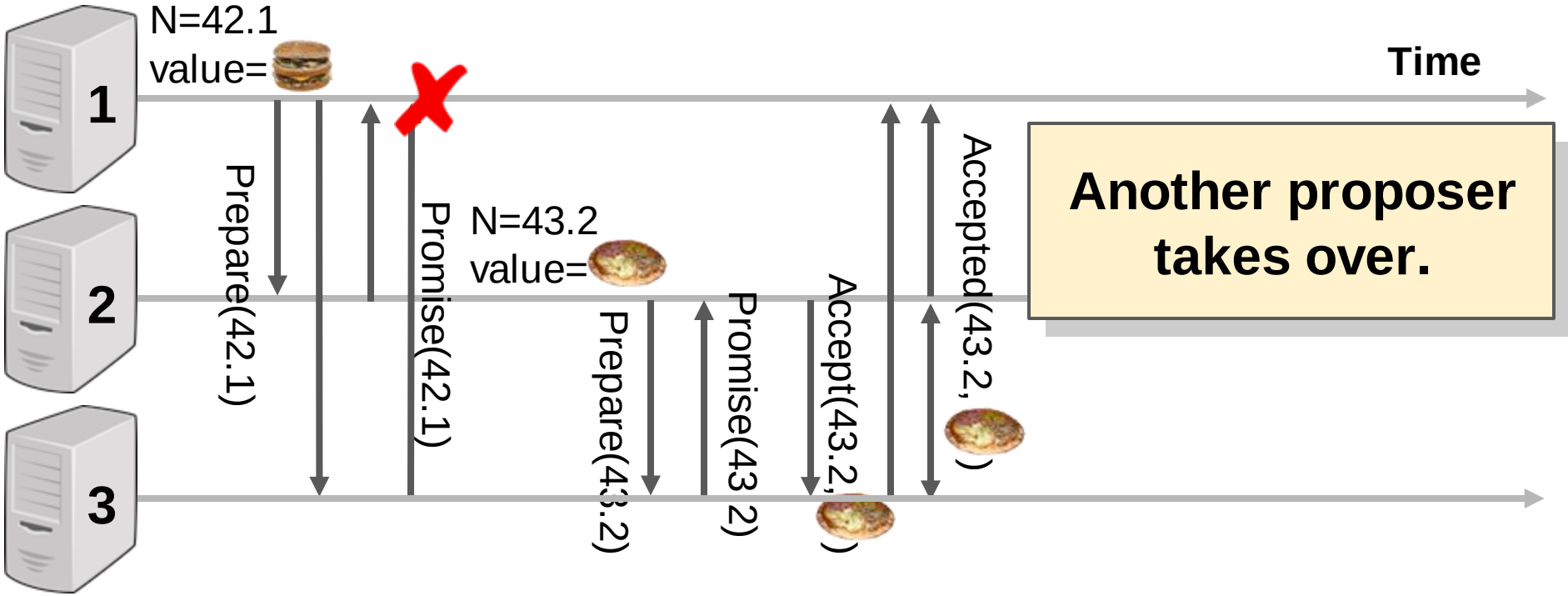
No problem as long as majority don't fail



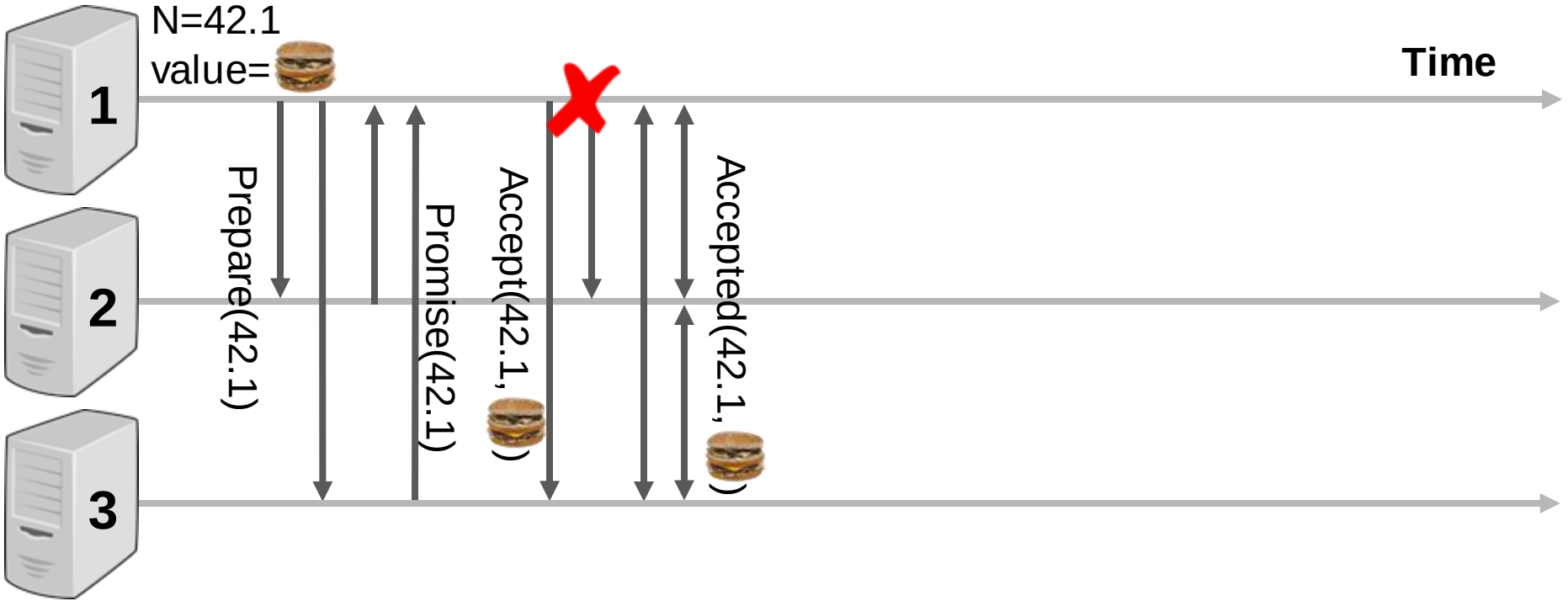
Failures: Proposer in Prepare Phase



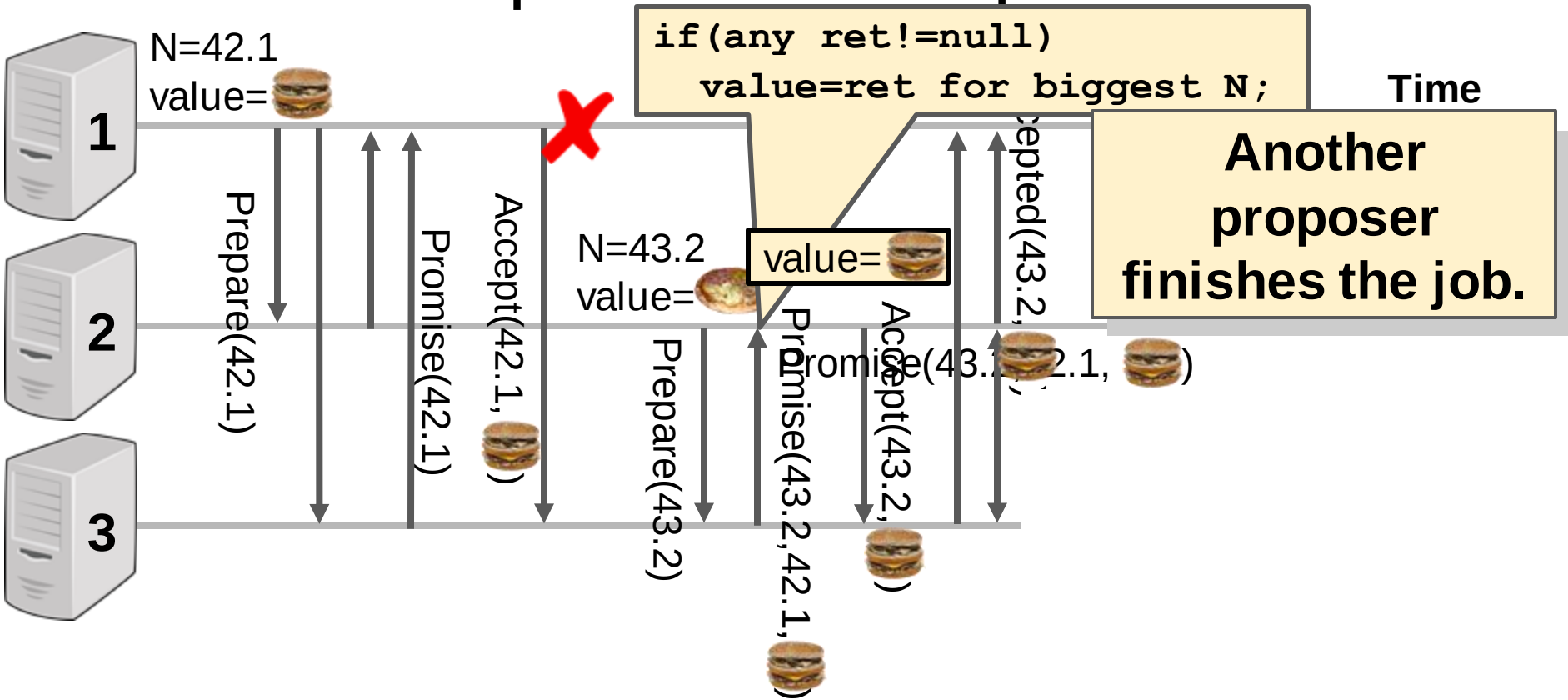
Failures: Proposer in Prepare Phase



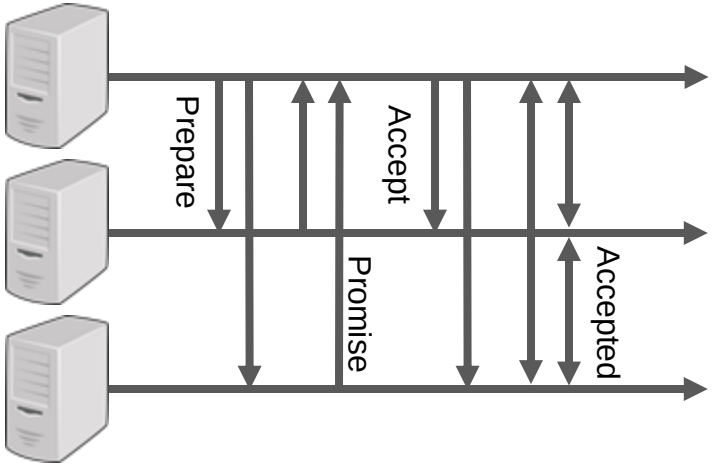
Failures: Proposer in Accept Phase



Failures: Proposer in Accept Phase



What could go wrong?



- One proposer
  - One or more acceptors fails
    - Still works as long as majority are up
  - Proposer fails in prepare phase
    - No-op; another proposer can make progress
  - Proposer fails in accept phase
    - Another proposer overwrites
    - Another proposer finishes the job
- Two or more simultaneous proposers
  - A bit more complex…
  - Can livelock, avoid with leader election

Paxos in the real world

- Creating a log of agreements
  - Multi-Paxos
- Adding and removing nodes from Paxos
  - Naming and cluster membership
- Testing and debugging is hard
- Byzantine failure
  - What happens if **N=∞**?

Need to implement?
Just use Raft.

Conclusions

- Consensus in a fail stop environment can be solved!
- Basic Paxos is not that hard
  - A full featured implementation is notoriously complex
- Learn Raft
  https://raft.github.io/