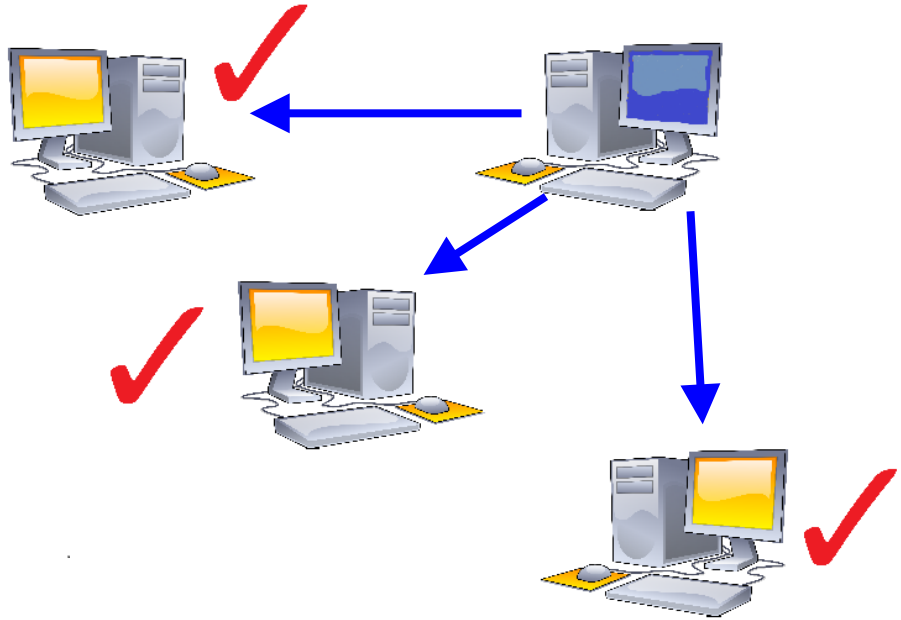
Investigating Consensus Algorithms

Anjali Saini
Mentor: Ling Ren

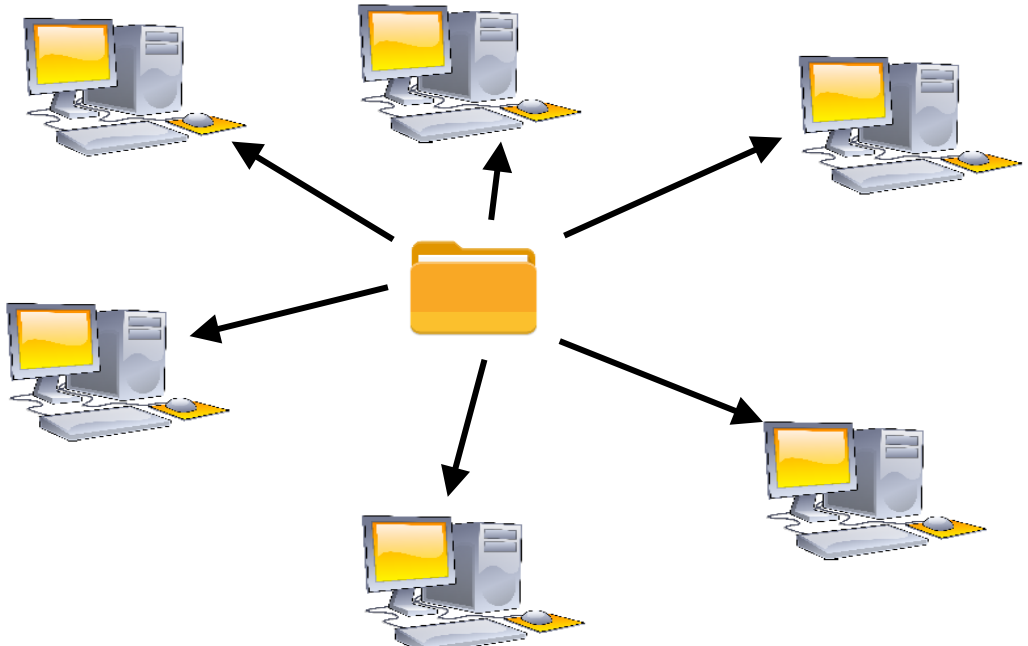
What is consensus?

- An important problem in distributed computing
- Processes agree on some data value
- Want to achieve global system reliability in the presence of faulty processes



Why do we want consensus?

- When storing files, must use multiple machines
- Making changes to that file requires all those machines to agree on the update



Consensus algorithms and fault tolerance

- Crash fault tolerance
  - Benign failures
  - Ex. Paxos
- Byzantine fault tolerance
  - Malicious failures
  - Ex. Practical Byzantine Fault Tolerance (PBFT)

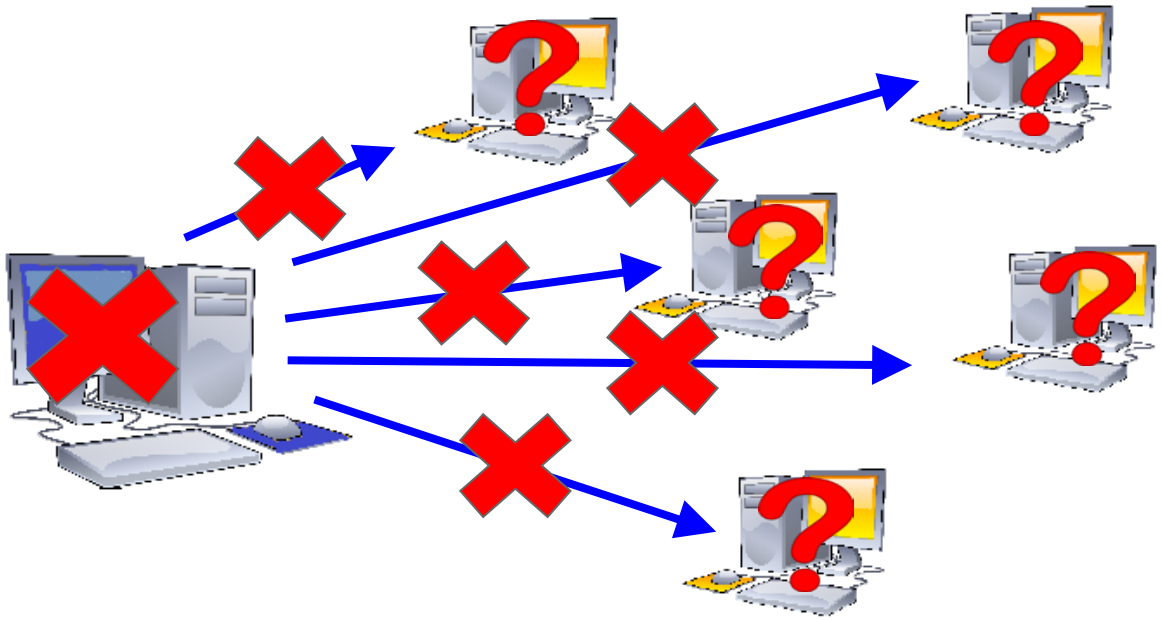
Paxos - Why is it useful?

The Paxos Algorithm

- A replica is a machine in the system that sends and receives messages
- The leader is the replica that makes proposals
- f is the maximum number of faulty replicas in the system
- Use a minimum of $2f+1$ replicas in the system to ensure consensus
- Goal: reach consensus

Approaches to Paxos

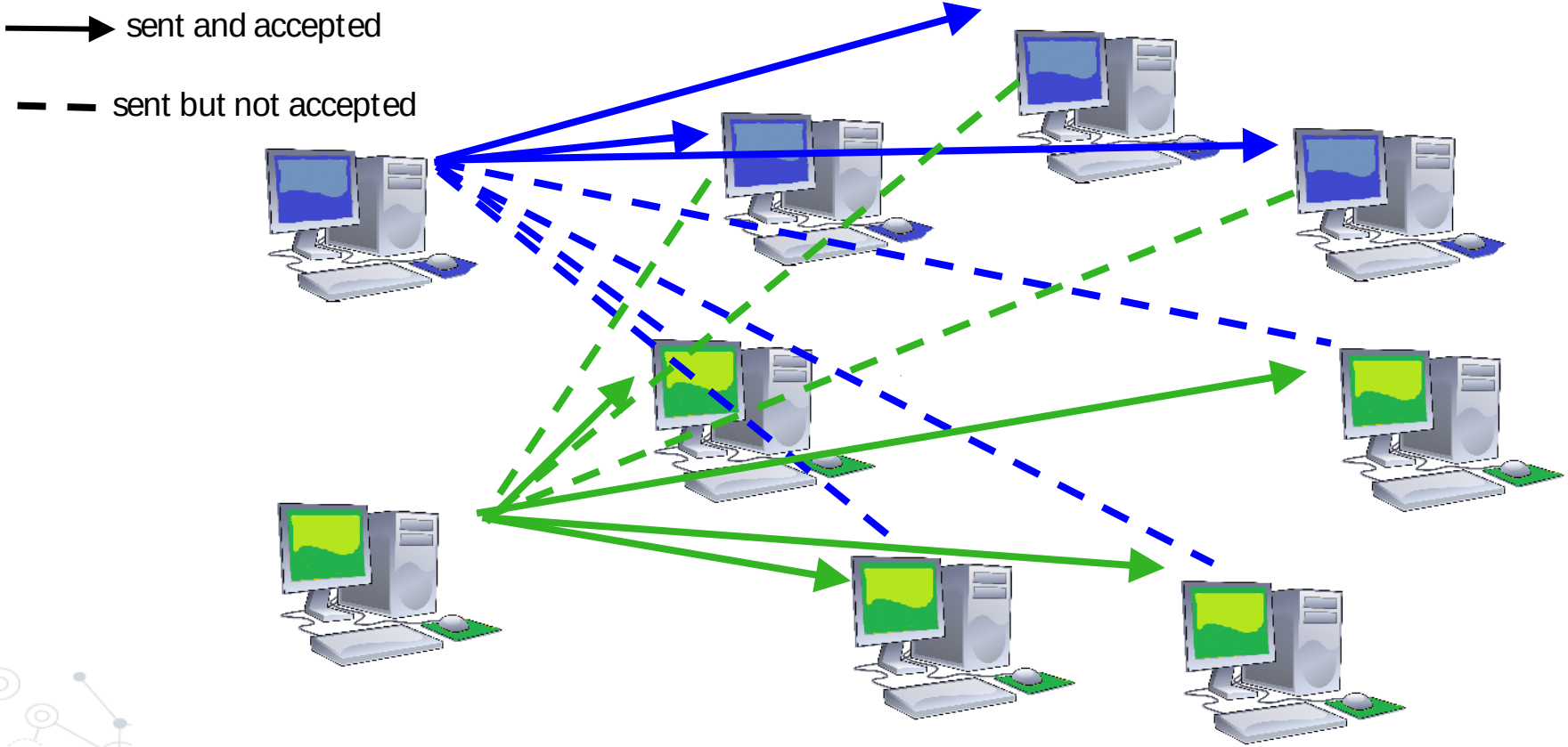
- Have one leader propose a value
  - Failure of that leader halts progress
  - Use multiple leaders



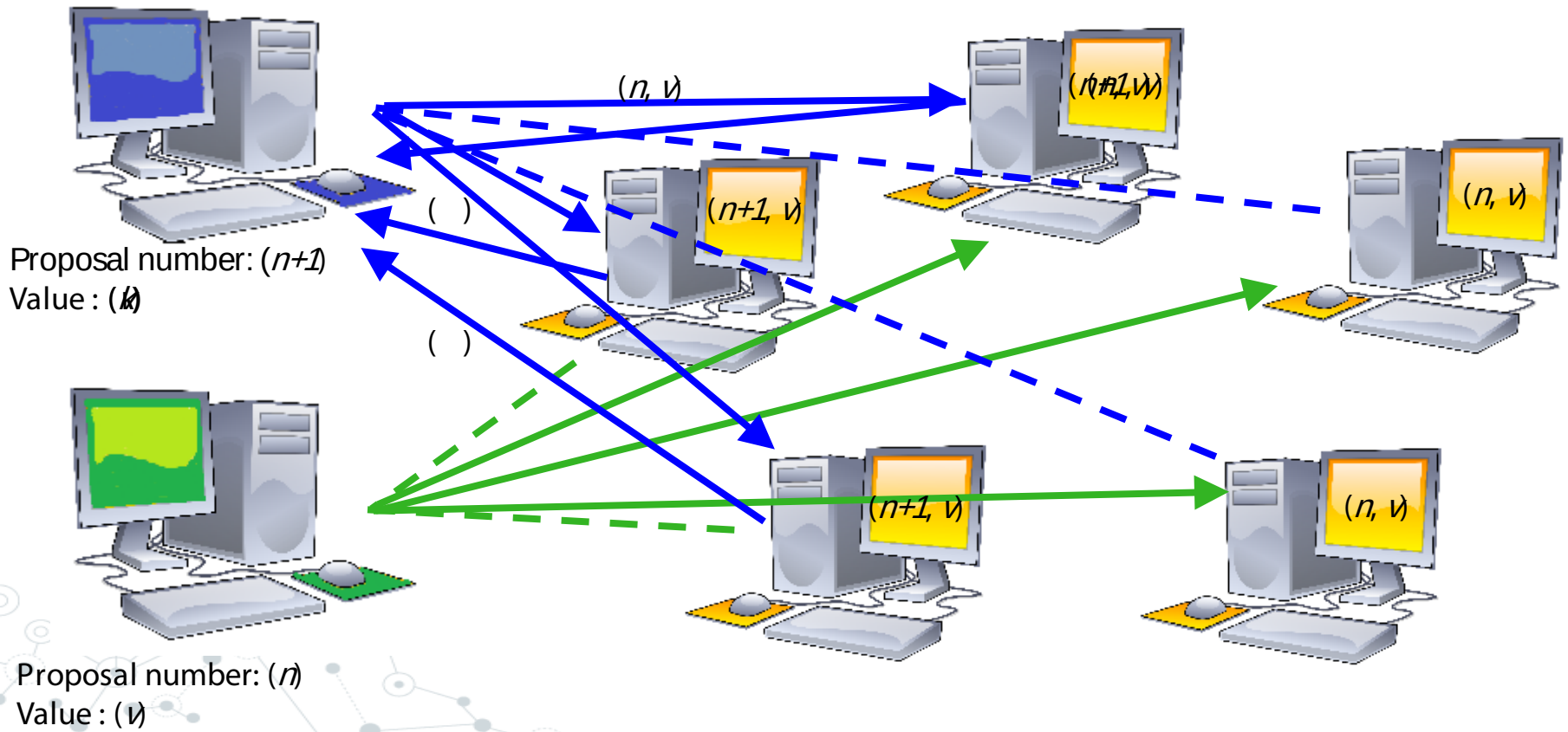
Approaches to Paxos

Now that there are multiple leaders to make proposals, let's have the replicas accept the first proposal that they receive.

Approaches to Paxos



The Paxos Algorithm



Our Paxos Implementation

- Separated into 3 steps
  - Phase 0
    - Leader collects information from at least $f+1$ replicas
  - Phase 1
    - Leader broadcasts its proposal
  - Phase 2
    - Replicas accept and then commit to the proposed value after receiving $f+1$ responses from other replicas

Consensus Algorithms - Uses/Applications

- Public key directory, Bitcoin, e-voting
- The Internet has many unknown users so there is a higher risk of malicious behavior
  - Hacking

Practical Byzantine Fault Tolerance (PBFT)

- Considers malicious attacks
- Uses $3f+1$ replicas to ensure consensus when at most f replicas are faulty

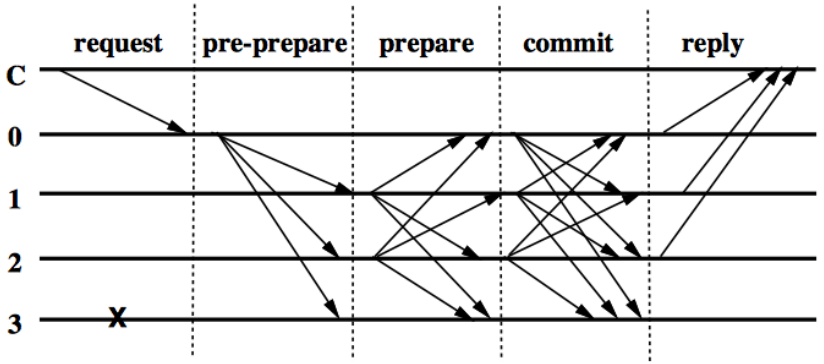


Figure 1: Normal Case Operation

Future Work

- Increase the scalability of consensus algorithms
  - Paxos is generally used is small scale, it has not been tested with more than approximately 10 replicas.
- Increase the scalability of PBFT
  - Move to an Internet scale

Acknowledgements

- My mentor, Ling Ren, for his guidance throughout the project
- Professor Srini Devadas for his helpful suggestions with the presentation
- The PRIMES program for providing this opportunity

Thank you! Any Questions?