Universitetet
i Stavanger

Quorum Selection for Byzantine Fault Tolerance

CREDENCE workshop 2019

Leander Jehl

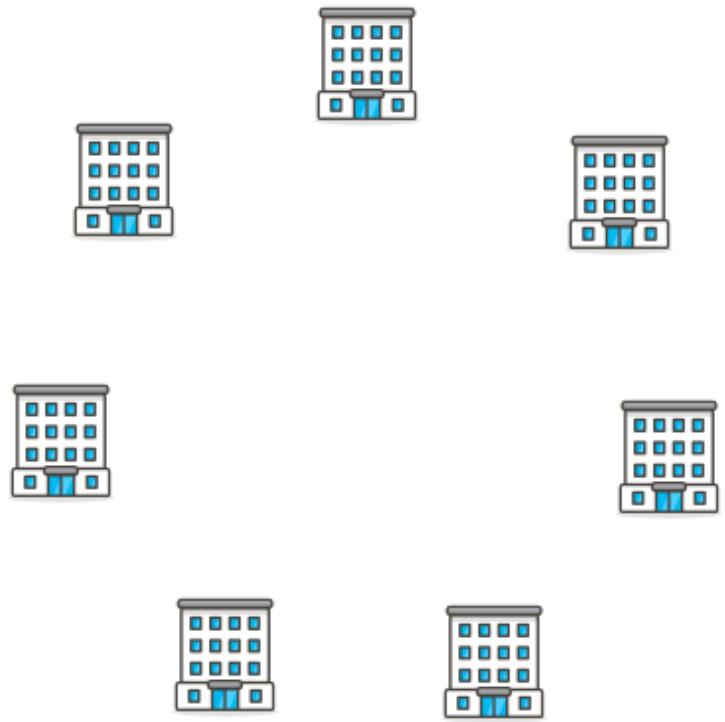
University of Stavanger, Norway

leander.jehl@uis.no

Permissioned BFT systems

Hyperledger Fabric, Tendermint,Symbiont, R3 Corda

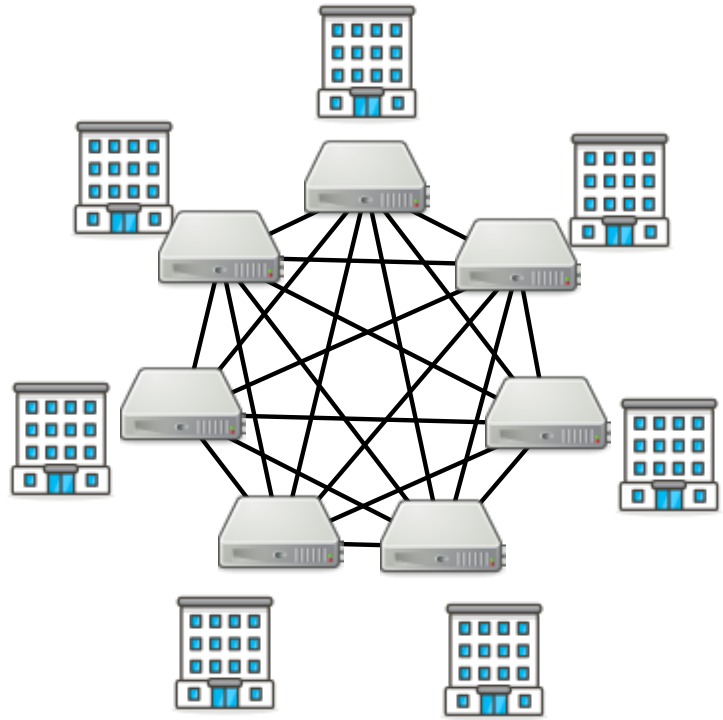
- shared between organizations
- conflicting interests



Permissioned BFT systems

Hyperledger Fabric, Tendermint,Symbiont, R3 Corda

- shared between organizations
- conflicting interests



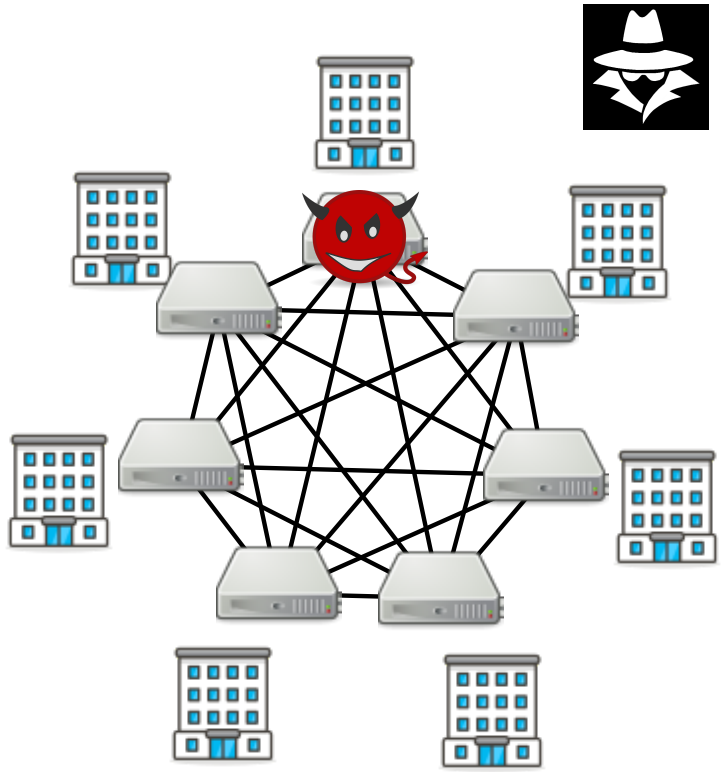
Permissioned BFT systems

Hyperledger Fabric, Tendermint,Symbiont, R3 Corda

- shared between organizations
- conflicting interests

failures

- caused without intent
- caused by an attacker



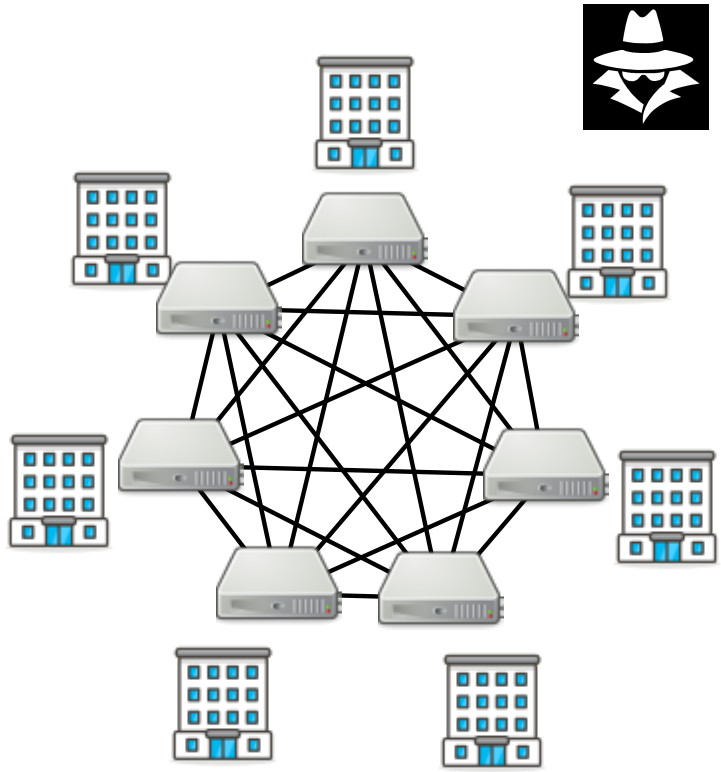
Permissioned BFT systems

Hyperledger Fabric, Tendermint,Symbiont, R3 Corda

- shared between organizations
- conflicting interests

failures

- caused without intent
- caused by an attacker
  - rejuvenation



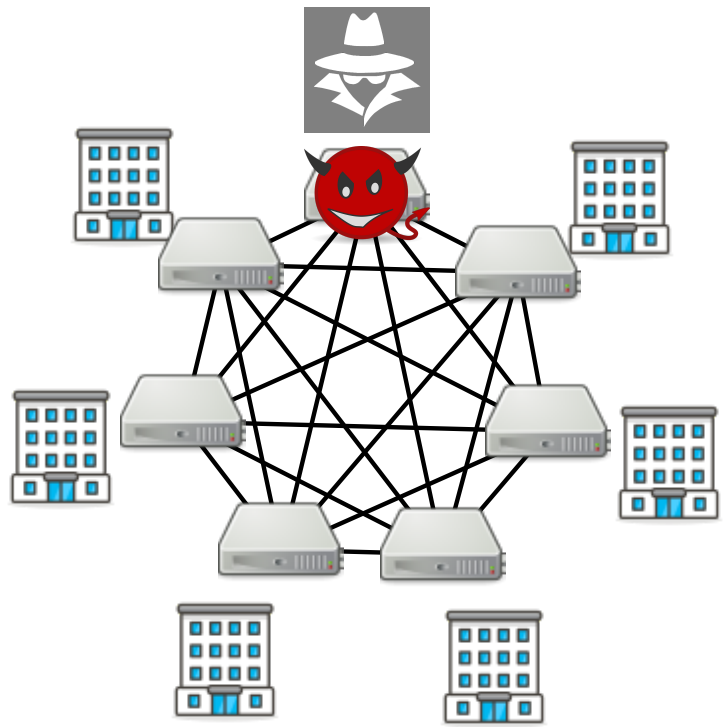
Permissioned BFT systems

Hyperledger Fabric, Tendermint,Symbiont, R3 Corda

- shared between organizations
- conflicting interests

failures

- caused without intent
- caused by an attacker
  - rejuvenation
- caused by a peer



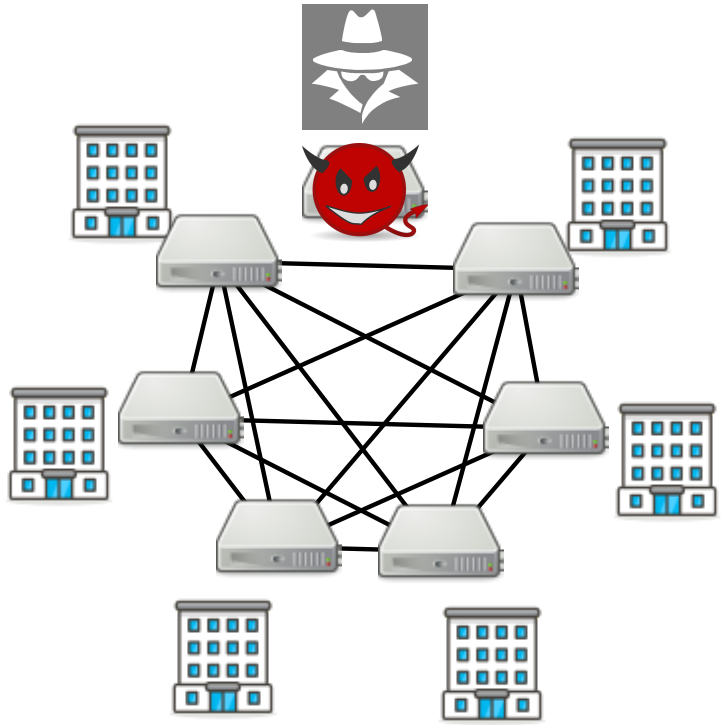
Permissioned BFT systems

Hyperledger Fabric, Tendermint,Symbiont, R3 Corda

- shared between organizations
- conflicting interests

failures

- caused without intent
- caused by an attacker
  - rejuvenation
- caused by a peer
  - exclude



Excluding replicas for performance

ReBFT: Optimization of PBFT [Distler et. al, TC'16]

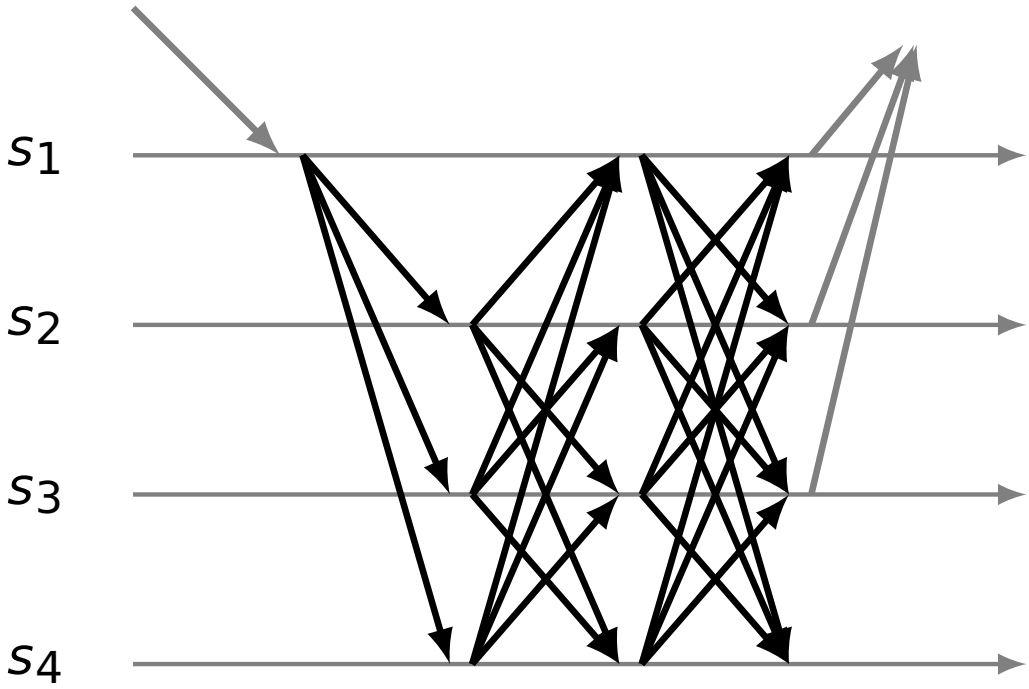


Figure: PBFT: Normal case messages

Excluding replicas for performance

ReBFT: Optimization of PBFT [Distler et. al, TC'16]

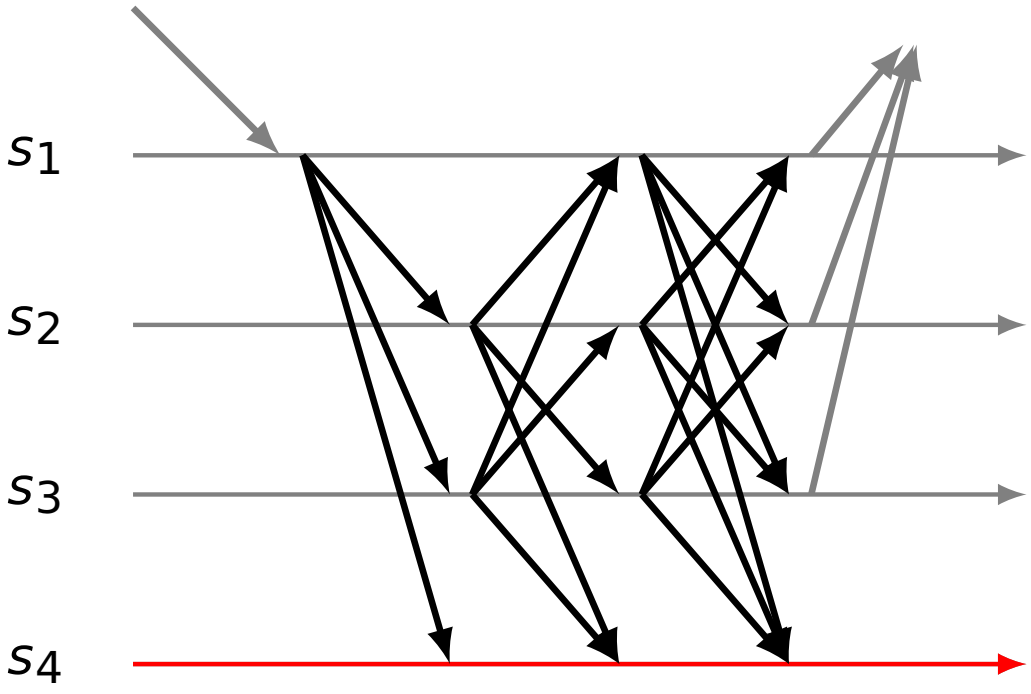


Figure: PBFT: Normal case, masks failure of s_4

Excluding replicas for performance

ReBFT: Optimization of PBFT [Distler et. al, TC'16]

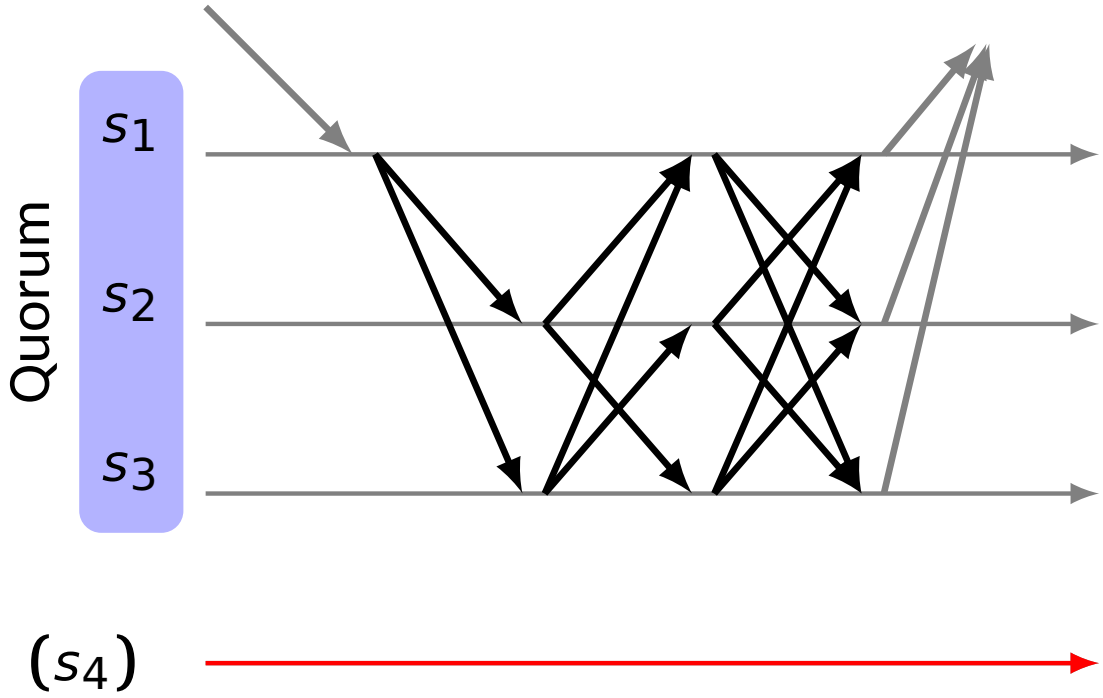


Figure: ReBFT: Throughput increased by 20%

Excluding replicas for performance

ReBFT: Optimization of PBFT [Distler et. al, TC'16]

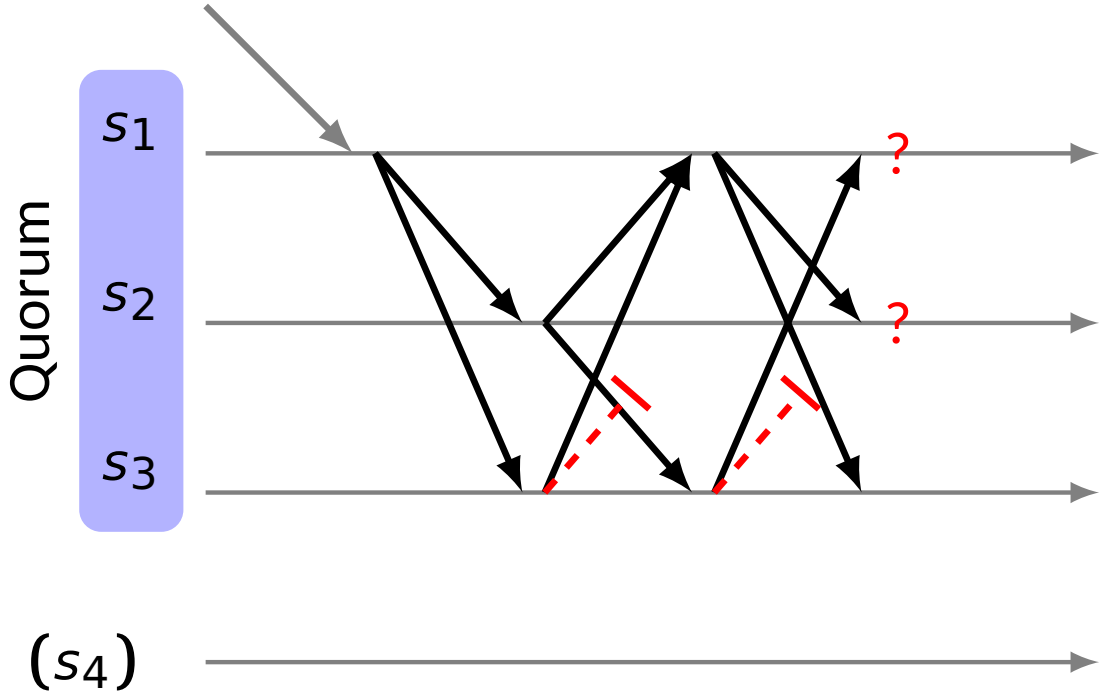


Figure: Omission stops progress

Excluding replicas for performance

ReBFT: Optimization of PBFT [Distler et. al, TC'16]

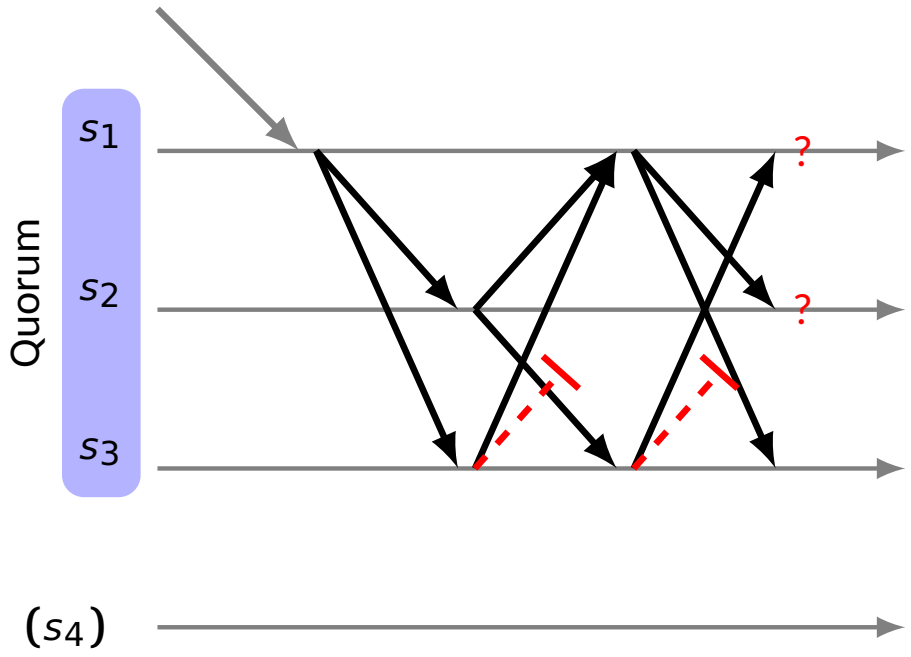


Figure: Omission stops progress

on failure

- fall back to PBFT

Excluding replicas for fault tolerance

XPaxos: [Liu et al., OSDI'16]

BFT with $2f + 1$ nodes in hybrid async/sync model

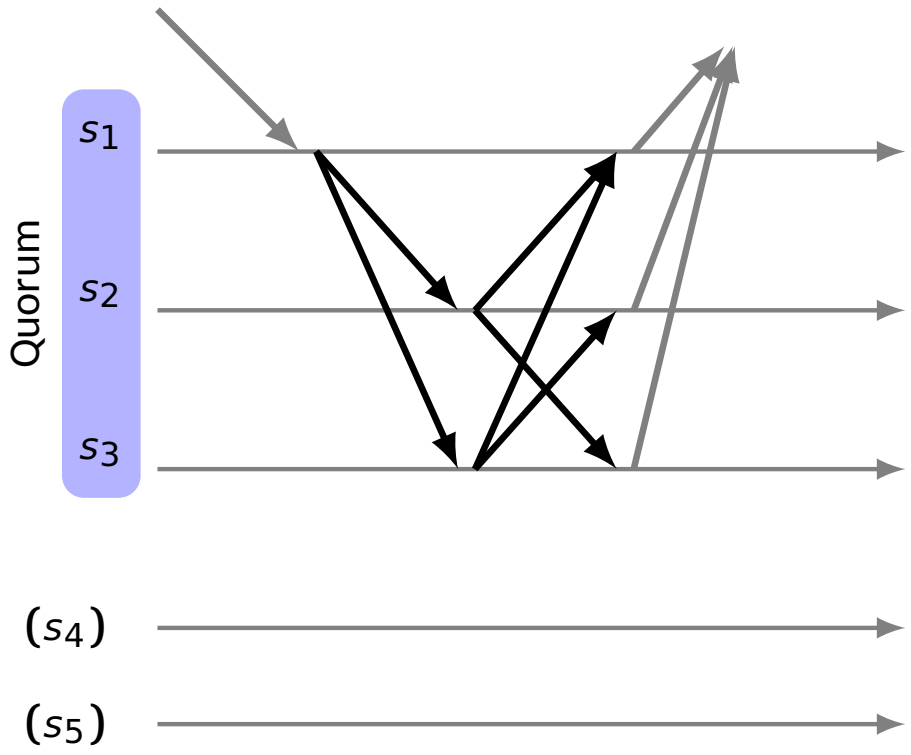
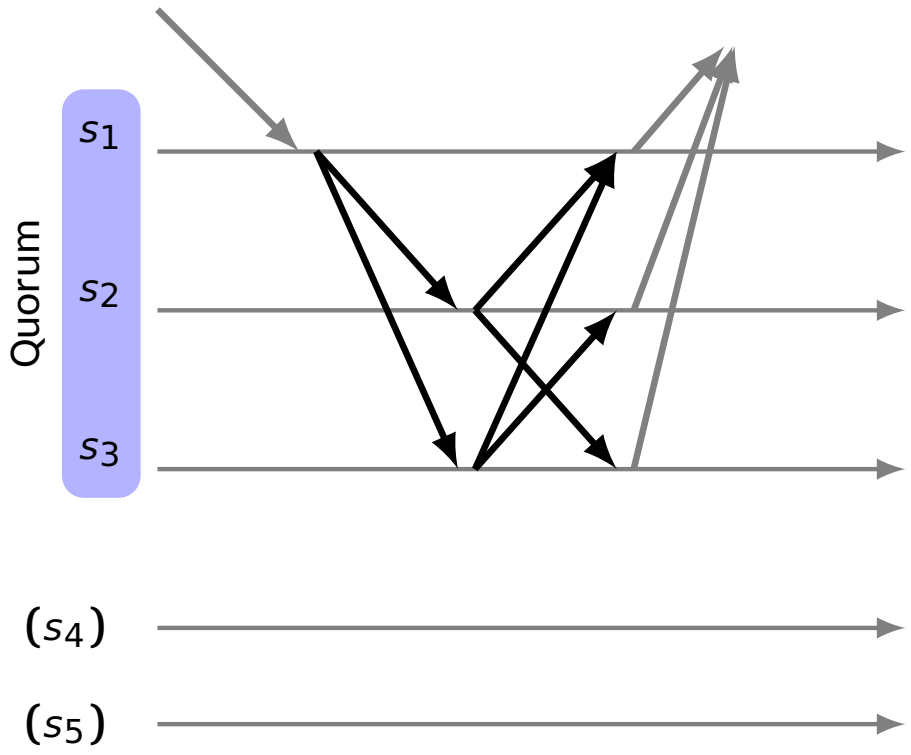


Figure: XPaxos with $f = 2$

Excluding replicas for fault tolerance

XPaxos: [Liu et al., OSDI'16]

BFT with $2f + 1$ nodes in hybrid async/sync model



on failure

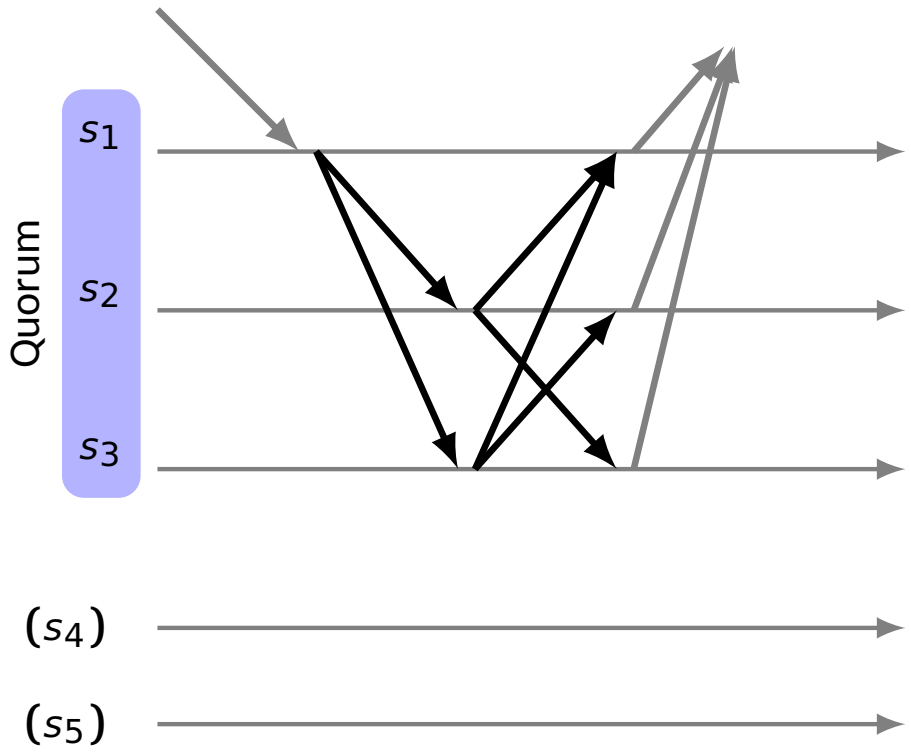
- try next quorum
- use round robin

Figure: XPaxos with $f = 2$

Excluding replicas for fault tolerance

XPaxos: [Liu et al., OSDI'16]

BFT with $2f + 1$ nodes in hybrid async/sync model



on failure

- try next quorum
- use round robin

  $\Omega(2^f)$ view changes

Figure: XPaxos with $f = 2$

Quorum-Selection

Architecture and algorithm to select a quorum containing correct/well behaved nodes.

Quorum-Selection: Architecture

System model

- $\Pi = \{s_1, s_2, \ldots\}$ nodes with $|\Pi| > 2f$
- up to f arbitrary failures
- asynchronous system with eventually accurate failure detector

Quorum-Selection: Architecture

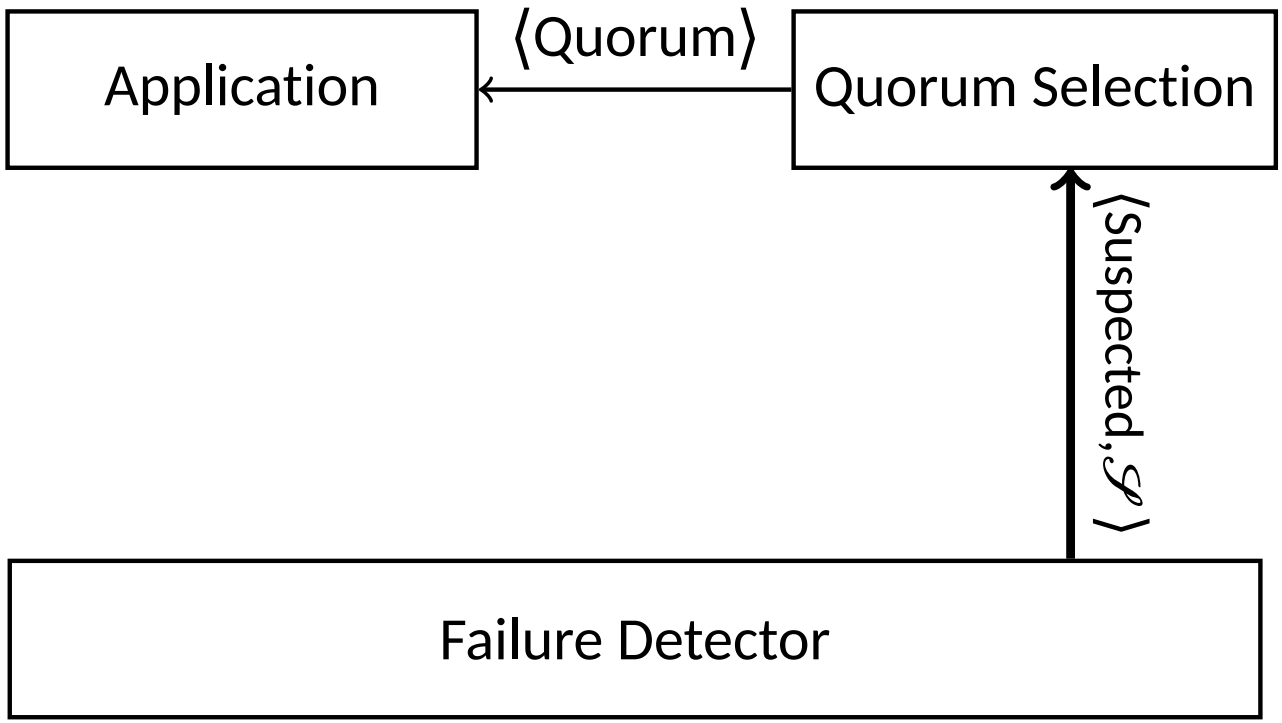


Figure: System components

Quorum-Selection: Architecture

- detection of failures depends on application

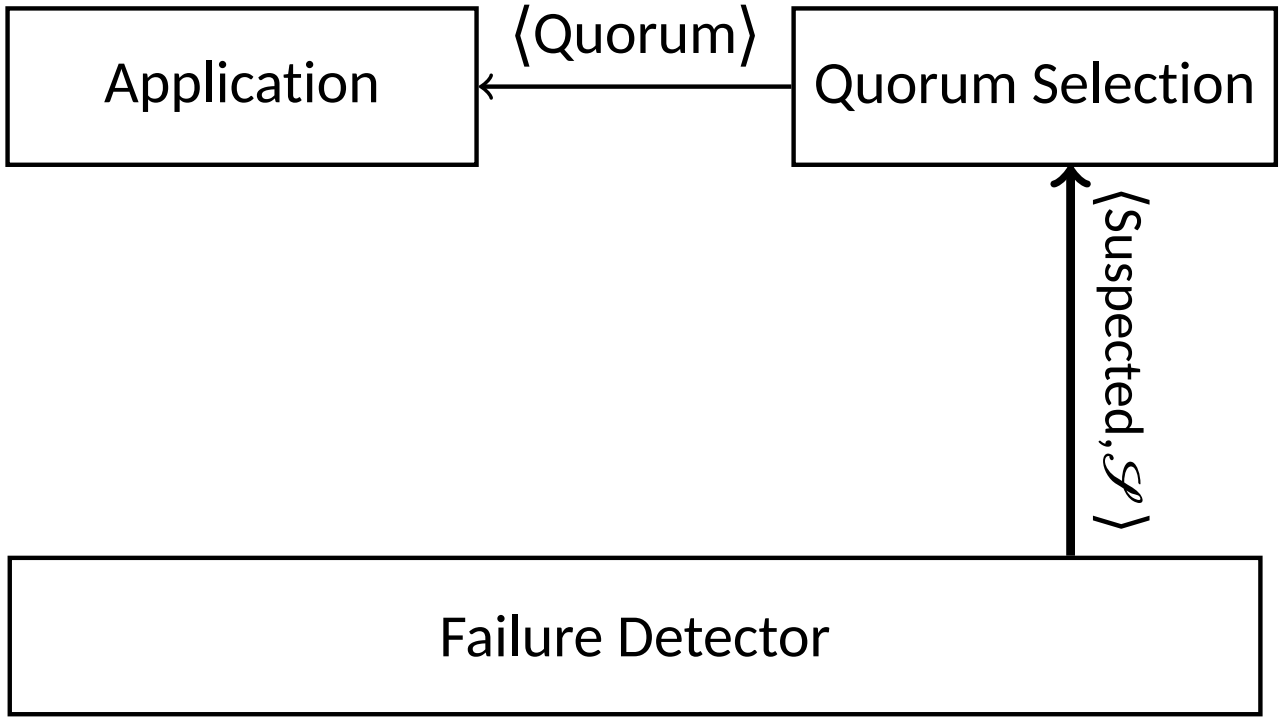
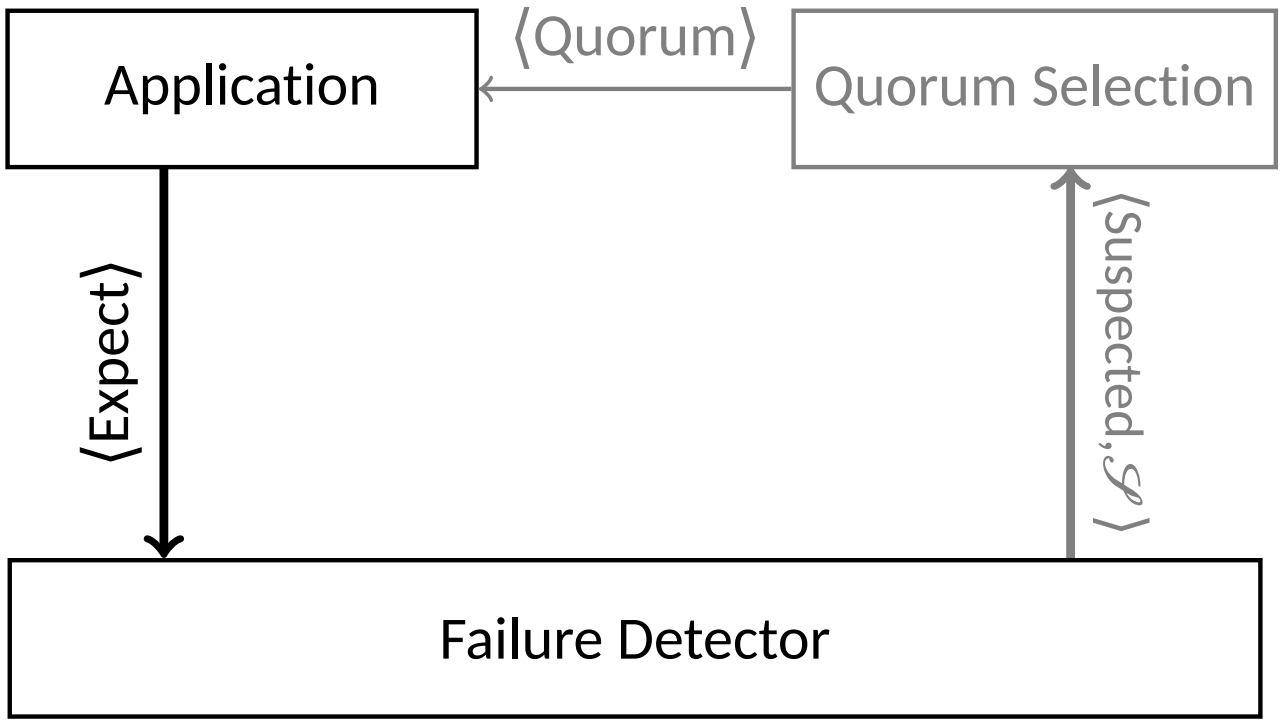


Figure: System components

Quorum-Selection: Architecture

Failure Detector

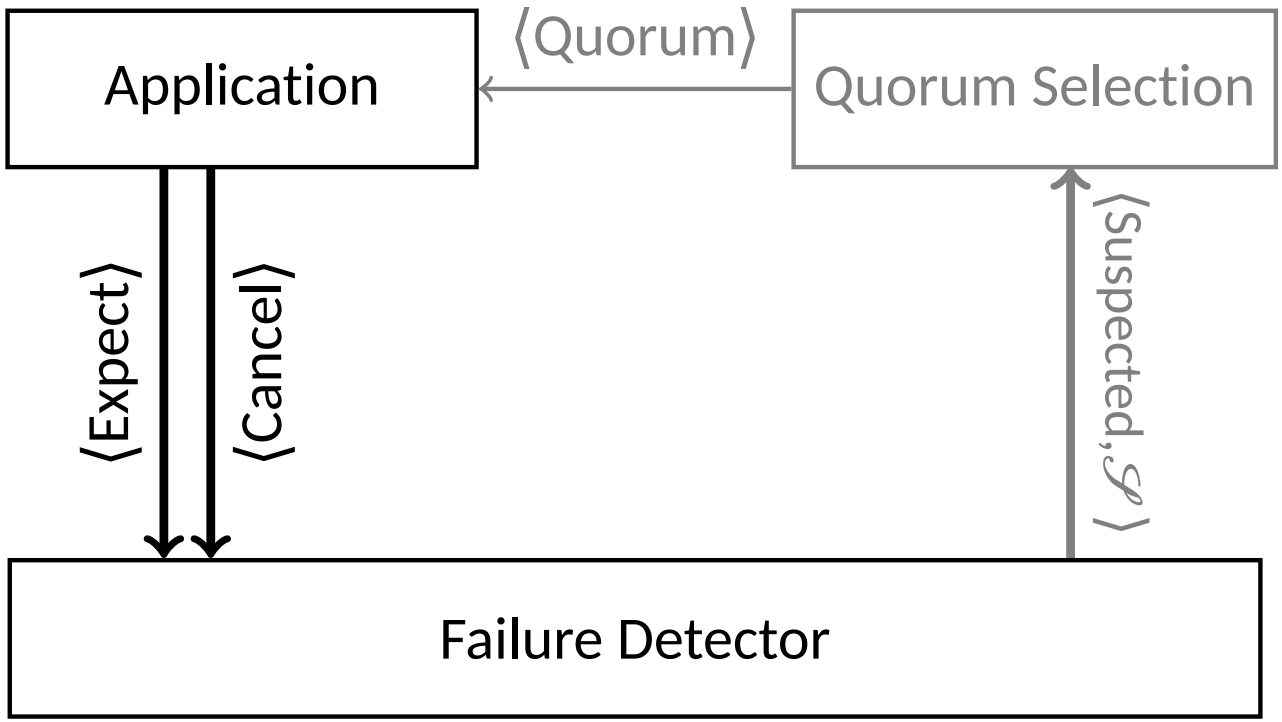
- detects omissions of **expected** messages



Quorum-Selection: Architecture

Failure Detector

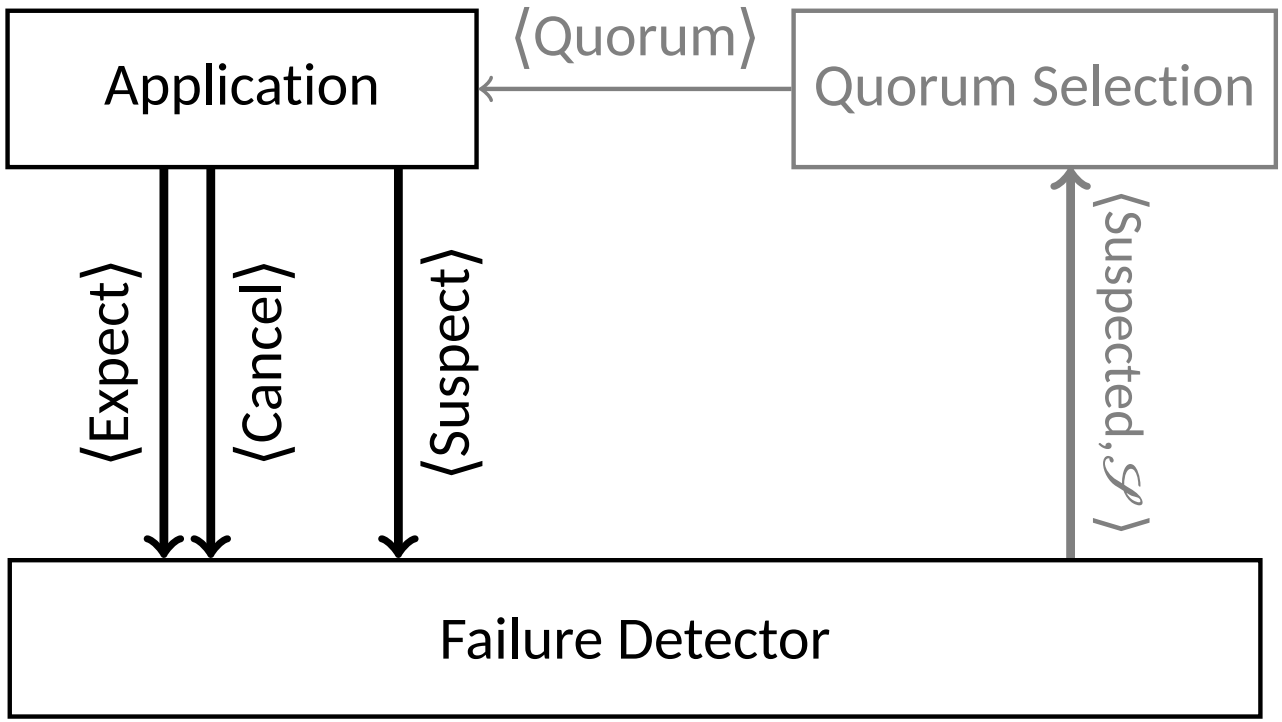
- detects omissions of **expected** messages



Quorum-Selection: Architecture

Failure Detector

- detects omissions of **expected** messages
- informed about commission failure/wrong messages



Quorum-Selection: Architecture

Failure Detector

- detects omissions of **expected** messages
- informed about commission failure/wrong messages

Failure Detector Assumptions

eventual strong accuracy

- eventually no suspicions between correct nodes

Quorum-Selection: Architecture

Failure Detector

- detects omissions of **expected** messages
- informed about commission failure/wrong messages

Failure Detector Assumptions

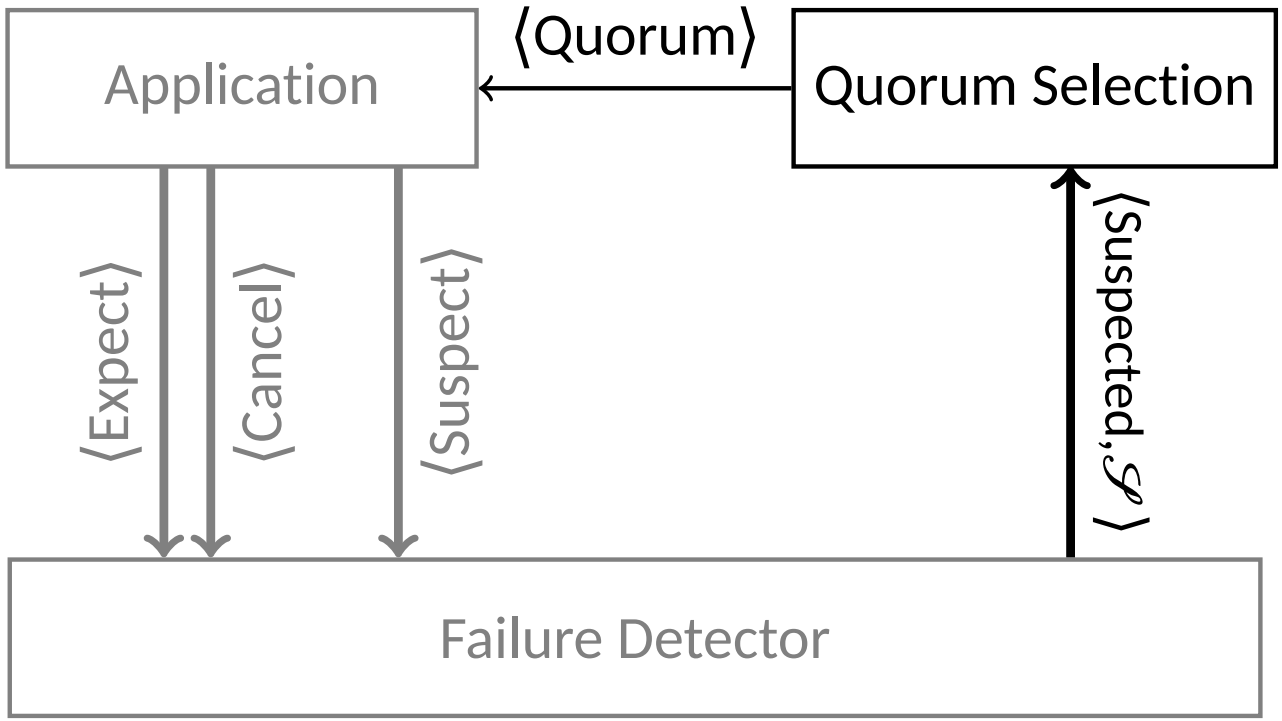
eventual strong accuracy

- eventually no suspicions between correct nodes

XPaxos example

see paper

Quorum-Selection



Quorum-Selection

Quorum-Selection Correctness

- correct processes eventually agree
- processes in the quorum do not suspect each other

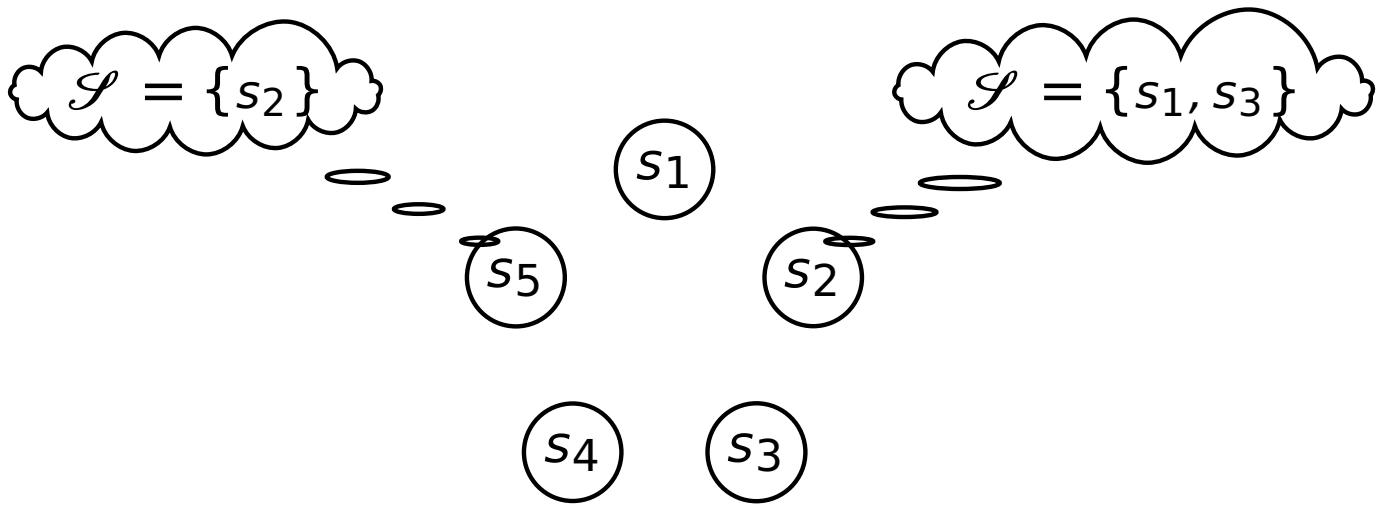
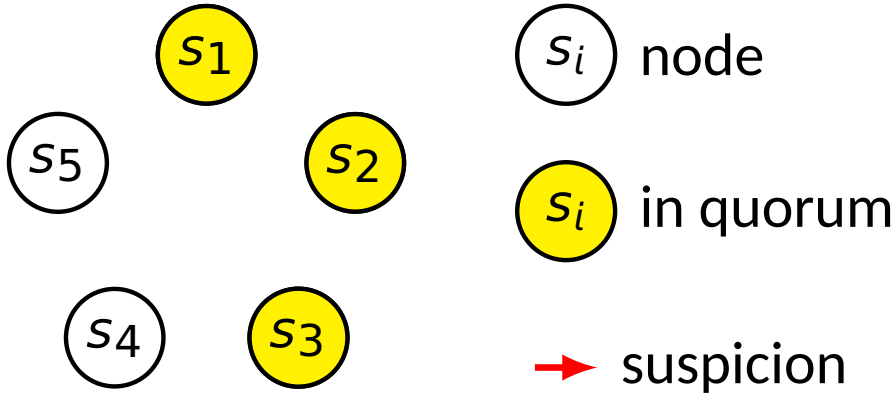


Figure: Nodes can disagree on suspicions

Quorum-Selection

Quorum-Selection Correctness

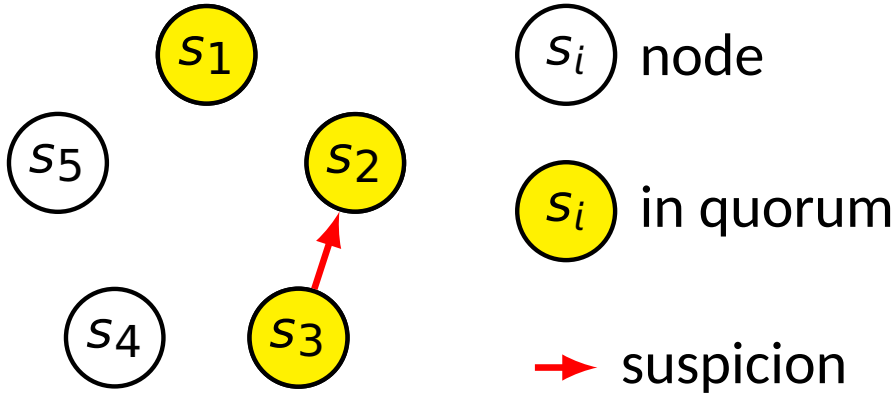
- correct processes eventually agree
- processes in the quorum do not suspect each other



Quorum-Selection

Quorum-Selection Correctness

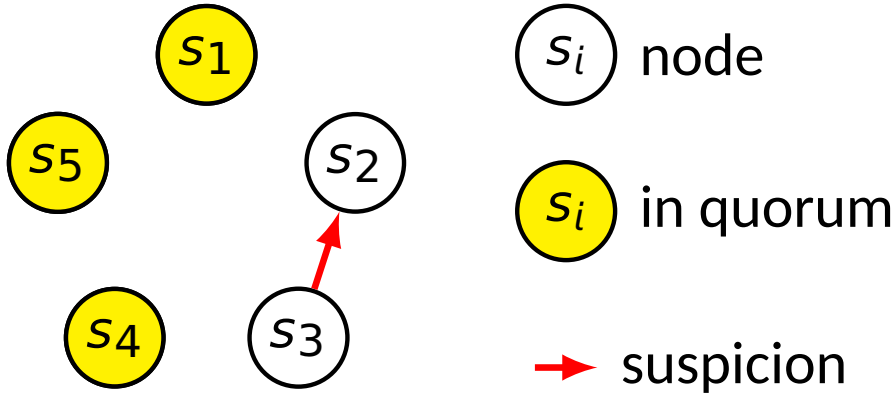
- correct processes eventually agree
- processes in the quorum do not suspect each other



Quorum-Selection

Quorum-Selection Correctness

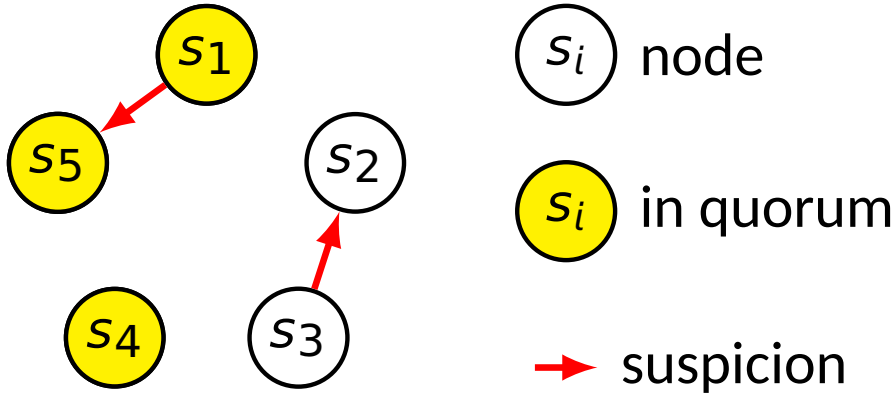
- correct processes eventually agree
- processes in the quorum do not suspect each other



Quorum-Selection

Quorum-Selection Correctness

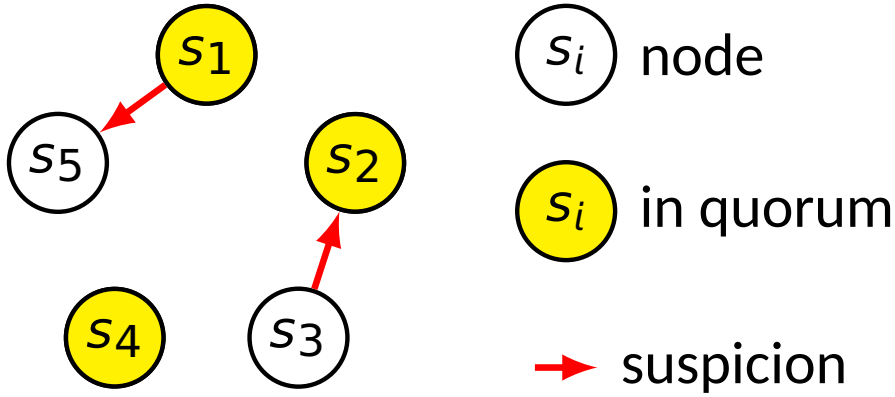
- correct processes eventually agree
- processes in the quorum do not suspect each other



Quorum-Selection

Quorum-Selection Correctness

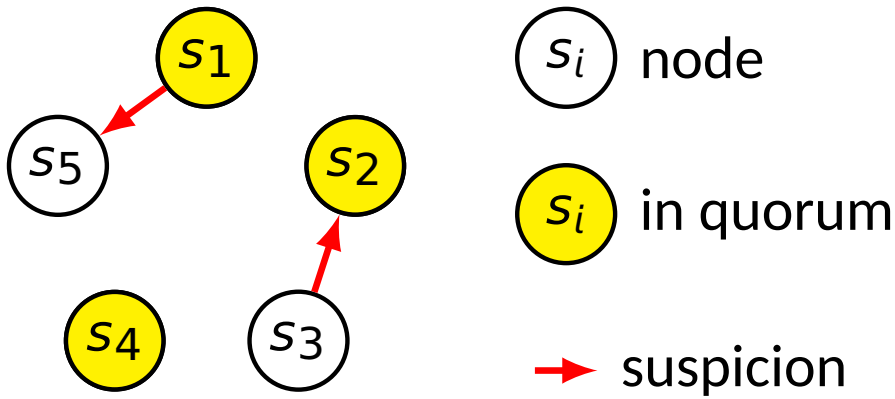
- correct processes eventually agree
- processes in the quorum do not suspect each other



Quorum-Selection

Quorum-Selection Correctness

- correct processes eventually agree
- processes in the quorum do not suspect each other

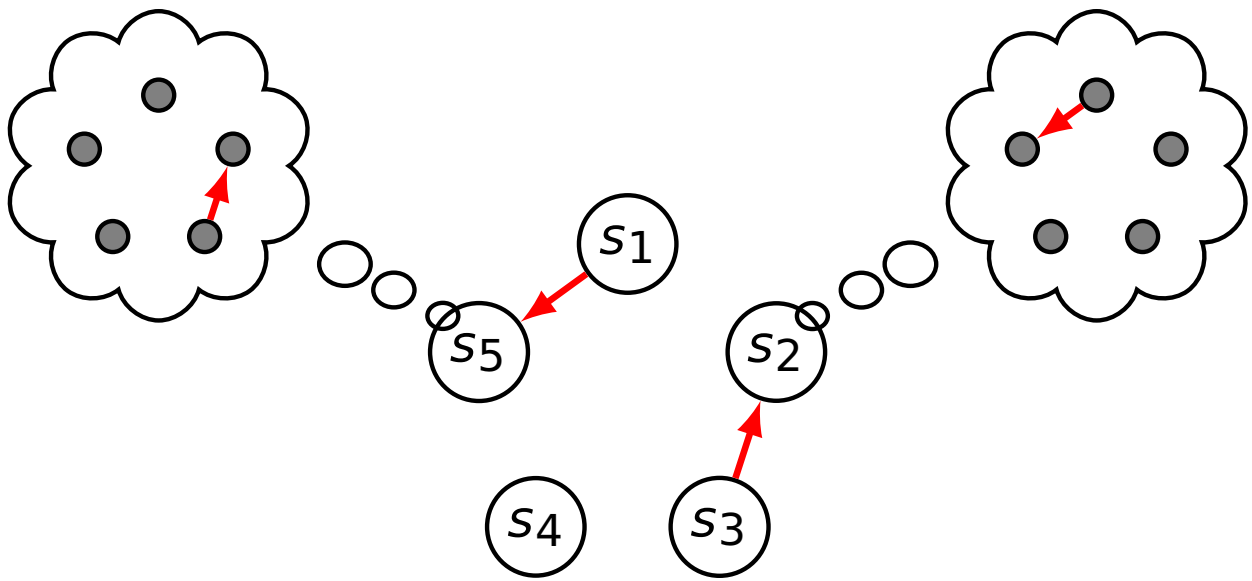


Metric

how many quorums issued, if failure detector is accurate

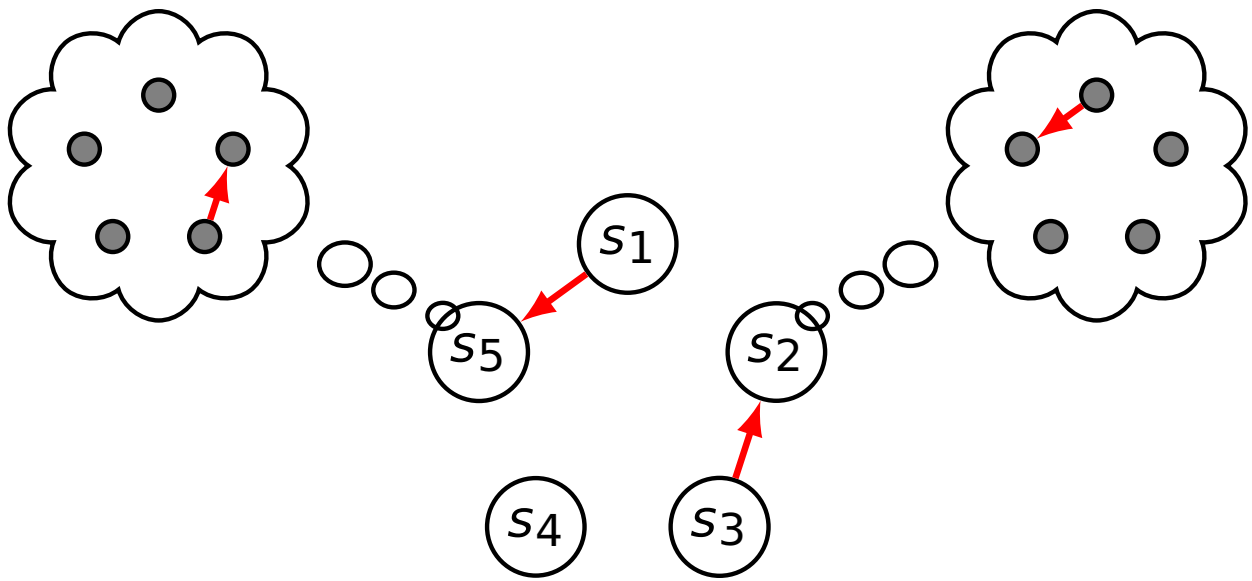
Suspect Graph

- all nodes collect suspicions



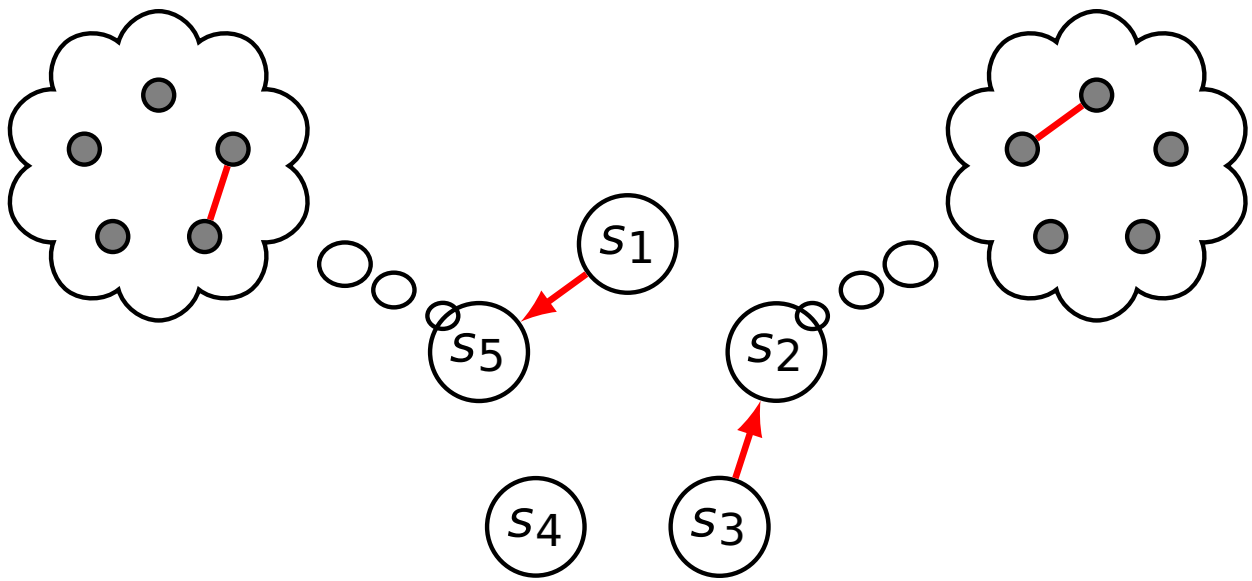
Suspect Graph

- all nodes collect suspicions
  - suspicions must be signed by suspecting node



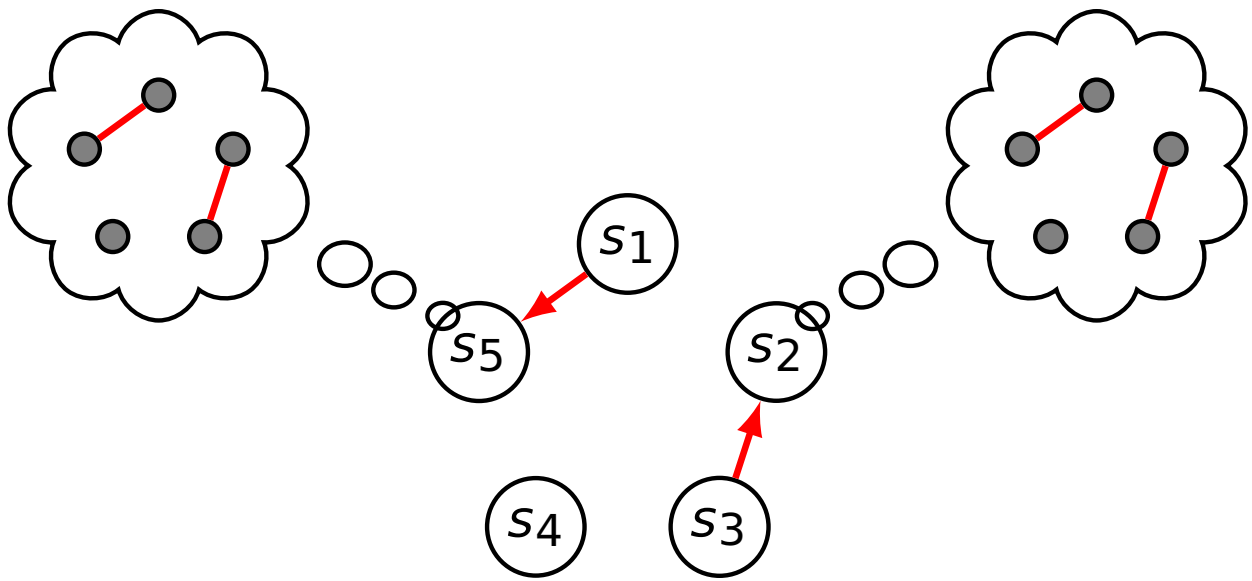
Suspect Graph

- all nodes collect suspicions
  - suspicions must be signed by suspecting node
- build simple graph



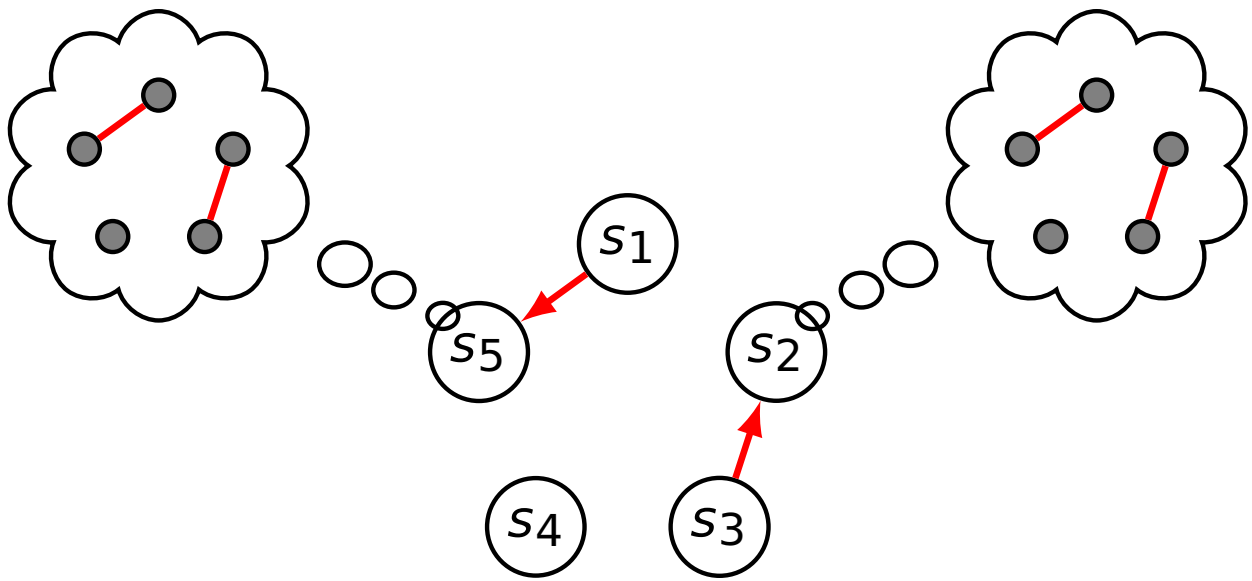
Suspect Graph

- all nodes collect suspicions
  - suspicions must be signed by suspecting node
- build simple graph
  - edges are not removed
  - correct nodes add the same edges in different order
  - eventually consistent



Suspect Graph

- all nodes collect suspicions
  - suspicions must be signed by suspecting node
- build simple graph
  - edges are not removed
  - correct nodes add the same edges in different order
  - eventually consistent
- find quorum as independent set of size $n - f$



Quorum-Selection false suspicions

Problem if the failure detector is not accurate, no independent set of size $n - f$ may exist

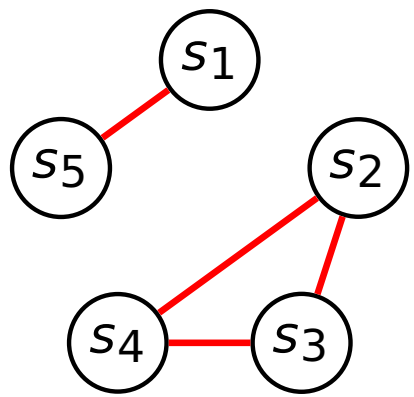


Figure: graph without independent set of size 3

Quorum-Selection false suspicions

Problem if the failure detector is not accurate, no independent set of size $n - f$ may exist

Solution

- assign epoch to suspicions
- when no quorum possible, increase epoch
- disregard suspicions from old epoch

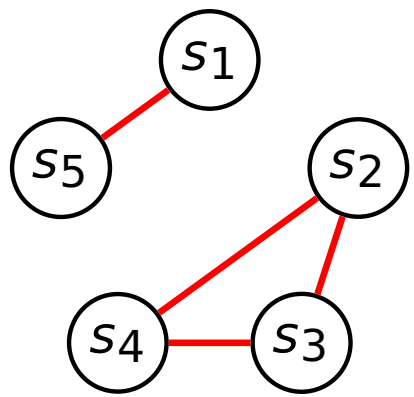


Figure: graph without independent set of size 3

Quorum-Selection

Metric

how many quorum issued, if failure detector is accurate

- we require $\mathcal{O}(f^2)$ quorums
- we proof a lower bound of $\Omega(f^2)$ quorums

Quorum-Selection

Metric

how many quorum issued, if failure detector is accurate

- we require $\mathcal{O}(f^2)$ quorums
- we proof a lower bound of $\Omega(f^2)$ quorums

Lower bound

Any deterministic algorithm requires at least $\binom{f+2}{2}$ quorum changes

Quorum-Selection

Metric

how many quorum issued, if failure detector is accurate

- we require $\mathcal{O}(f^2)$ quorums
- we proof a lower bound of $\Omega(f^2)$ quorums

Lower bound

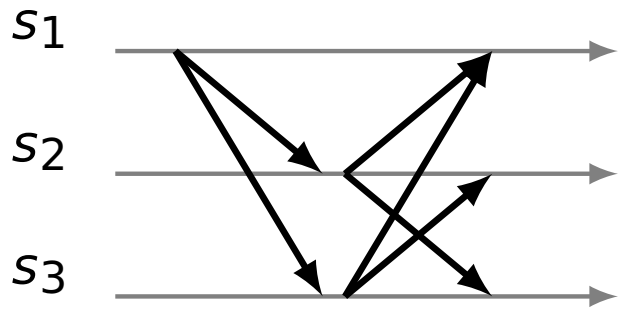
Any deterministic algorithm requires at least $\binom{f+2}{2}$ quorum changes

Idea concentrate suspicions on 2 correct nodes

Quorum-Selection Variations

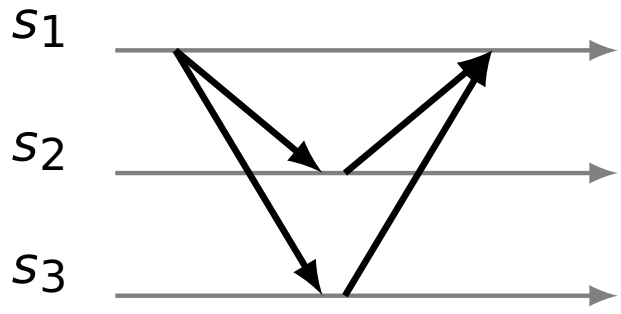
All-to-all algorithms

need to react on any suspicion within quorum



Leader based algorithms

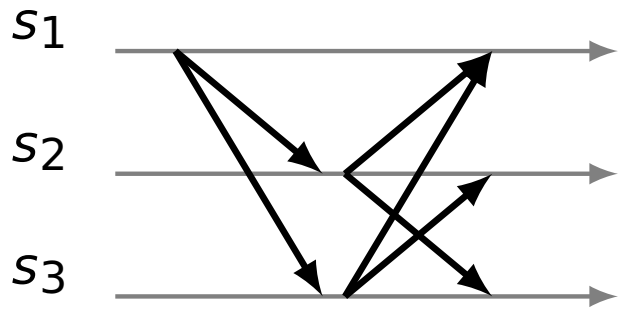
ignore suspicions between followers



Quorum-Selection Variations

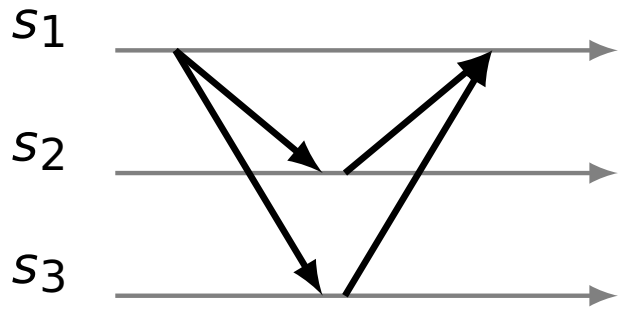
All-to-all algorithms

need to react on any suspicion within quorum



Leader based algorithms

ignore suspicions between followers



Follower-Selection

- assume $|\Pi| > 3f$
- only $\mathcal{O}(f)$ quorums

Follower-Selection

Idea

- let leader select followers
- every leader only gets one try

Follower-Selection

Idea

- let leader select followers
- every leader only gets one try

- use failure detector to suspect misbehaving leader

Follower-Selection

Idea

- let leader select followers
- every leader only gets one try

- use failure detector to suspect misbehaving leader

---

at most $6f$ quorums with accurate failure detector

Quorum-Selection

- architecture
- eventual consistent suspect graph
- quorum as independent set in $\Theta(f^2)$ changes

Quorum-Selection

- architecture
- eventual consistent suspect graph
- quorum as independent set in $\Theta(f^2)$ changes

Follower-Selection

- no all-to-all communication
- $|\Pi| > 3f$
- only $\mathcal{O}(f)$ changes

Quorum-Selection

- architecture
- eventual consistent suspect graph
- quorum as independent set in $\Theta(f^2)$ changes

Follower-Selection

- no all-to-all communication
- $|\Pi| > 3f$
- only $\mathcal{O}(f)$ changes

Open Questions

- other communication patterns
- Follower-Selection with $|\Pi| = 2f + 1$

Questions?

Quorum-Selection

Metric

how many quorum issued, if failure detector is accurate

- we require $\mathcal{O}(f^2)$ quorums
- we proof a lower bound of $\Omega(f^2)$ quorums

Quorum-Selection

Metric

how many quorum issued, if failure detector is accurate

- we require $\mathcal{O}(f^2)$ quorums
- we proof a lower bound of $\Omega(f^2)$ quorums

Lower bound

Any deterministic algorithm requires at least $\binom{f+2}{2}$ quorum changes

Quorum-Selection

Metric

how many quorum issued, if failure detector is accurate

- we require $\mathcal{O}(f^2)$ quorums
- we proof a lower bound of $\Omega(f^2)$ quorums

Lower bound

Any deterministic algorithm requires at least $\binom{f+2}{2}$ quorum changes

Assumption

- faulty node may suspect anybody
- faulty node may cause to be suspected by anybody

Quorum-Selection

Metric

how many quorum issued, if failure detector is accurate

- we require $\mathcal{O}(f^2)$ quorums
- we proof a lower bound of $\Omega(f^2)$ quorums

Lower bound

Any deterministic algorithm requires at least $\binom{f+2}{2}$ quorum changes

Assumption

- faulty node may suspect anybody
- faulty node may cause to be suspected by anybody

Idea concentrate suspicions on 2 correct nodes

Follower-Selection

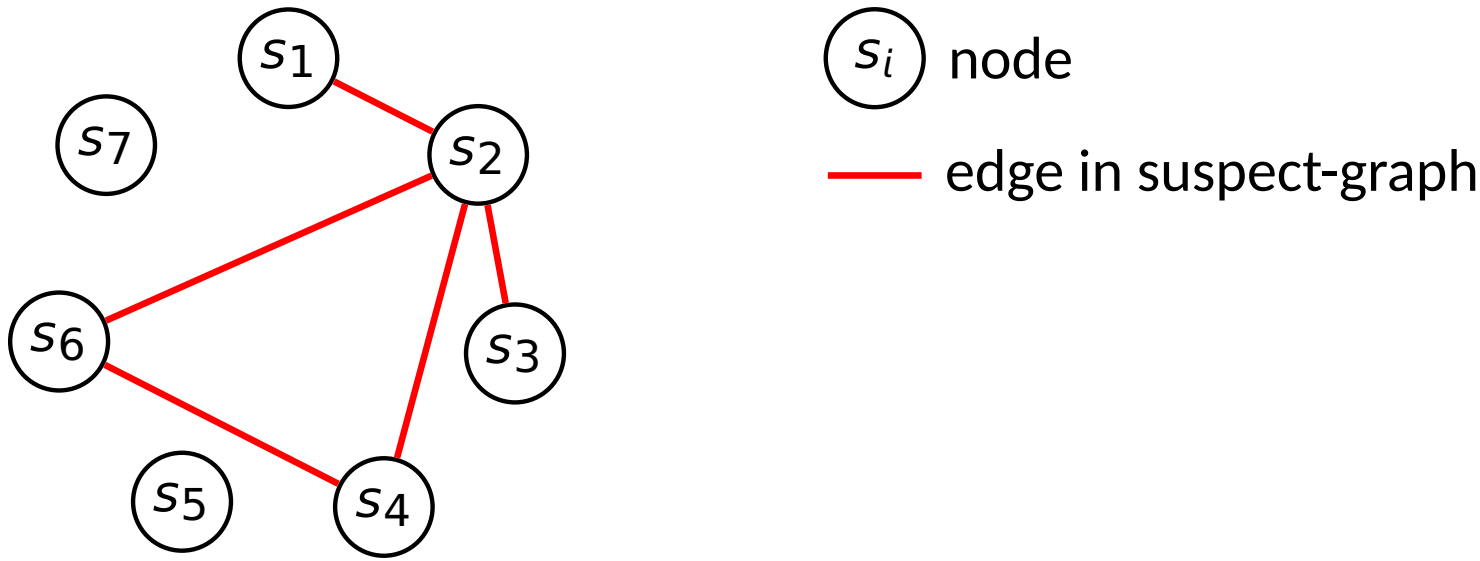


Figure: example graph with $f = 2$

Follower-Selection

- find subgraph L, acyclic with maximum degree 2

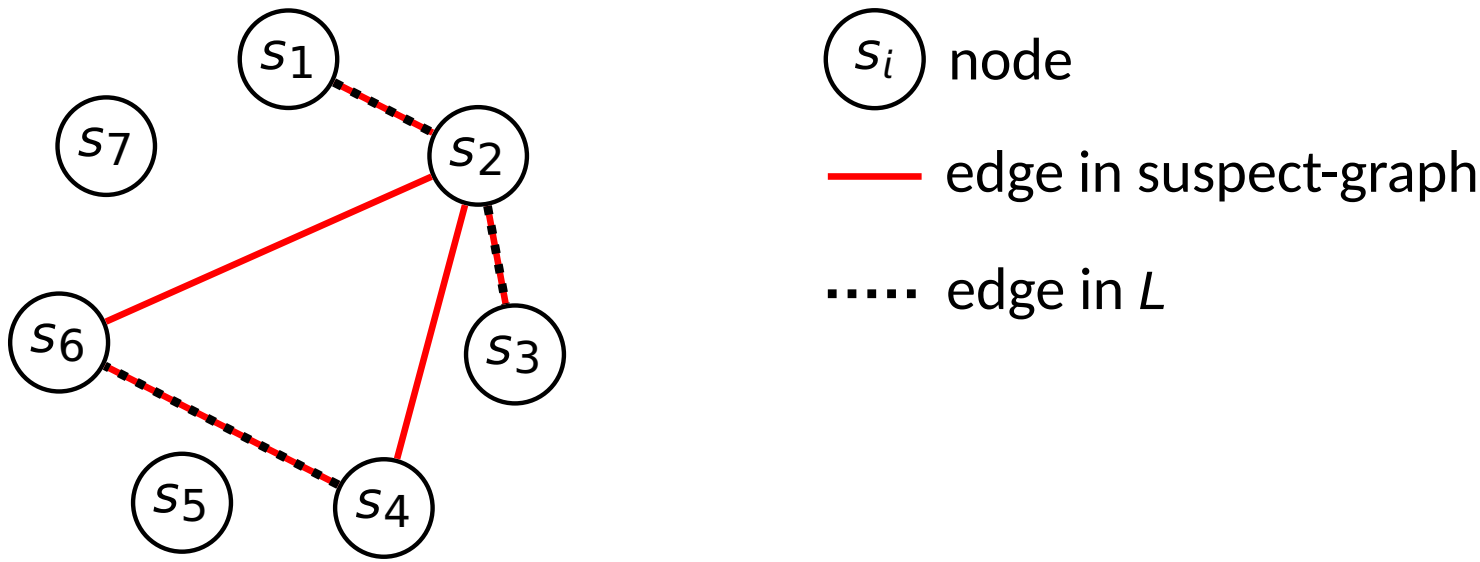


Figure: example graph with $f = 2$

Follower-Selection

- find subgraph L, acyclic with maximum degree 2
- select a leader
  - node with degree 0 in L

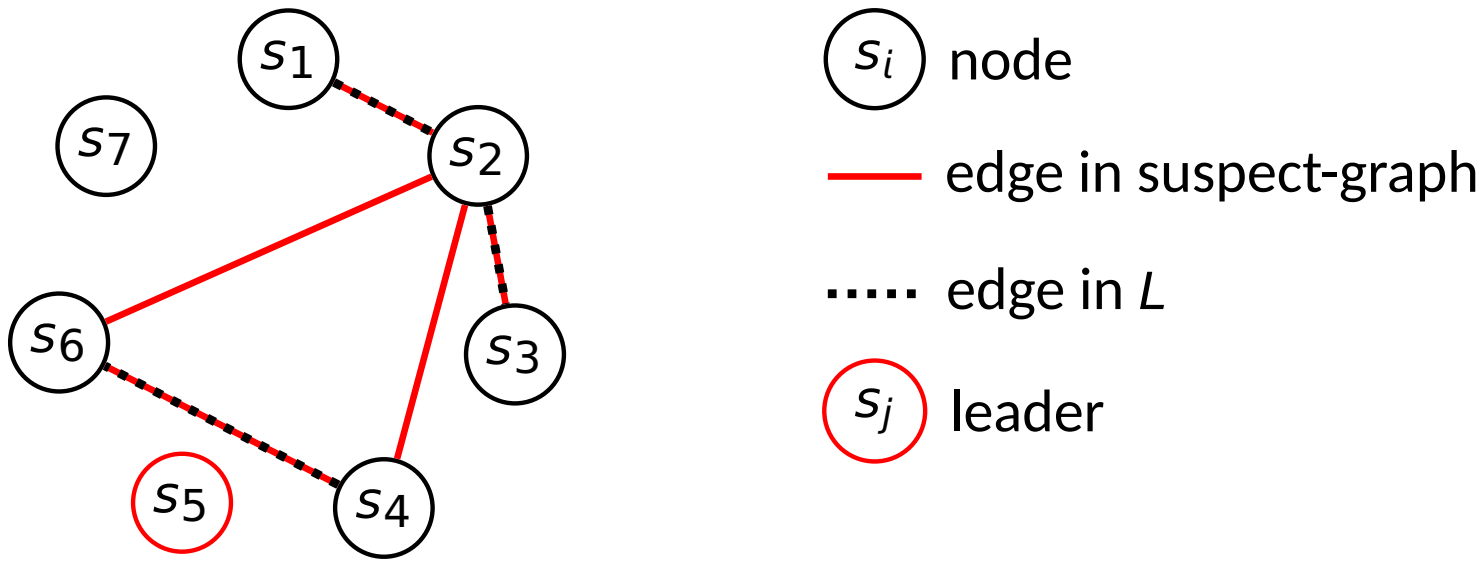


Figure: example graph with $f = 2$

Follower-Selection

- find subgraph L, acyclic with maximum degree 2
- select a leader
  - node with degree 0 in L
- leader selects followers of degree 0 or 1

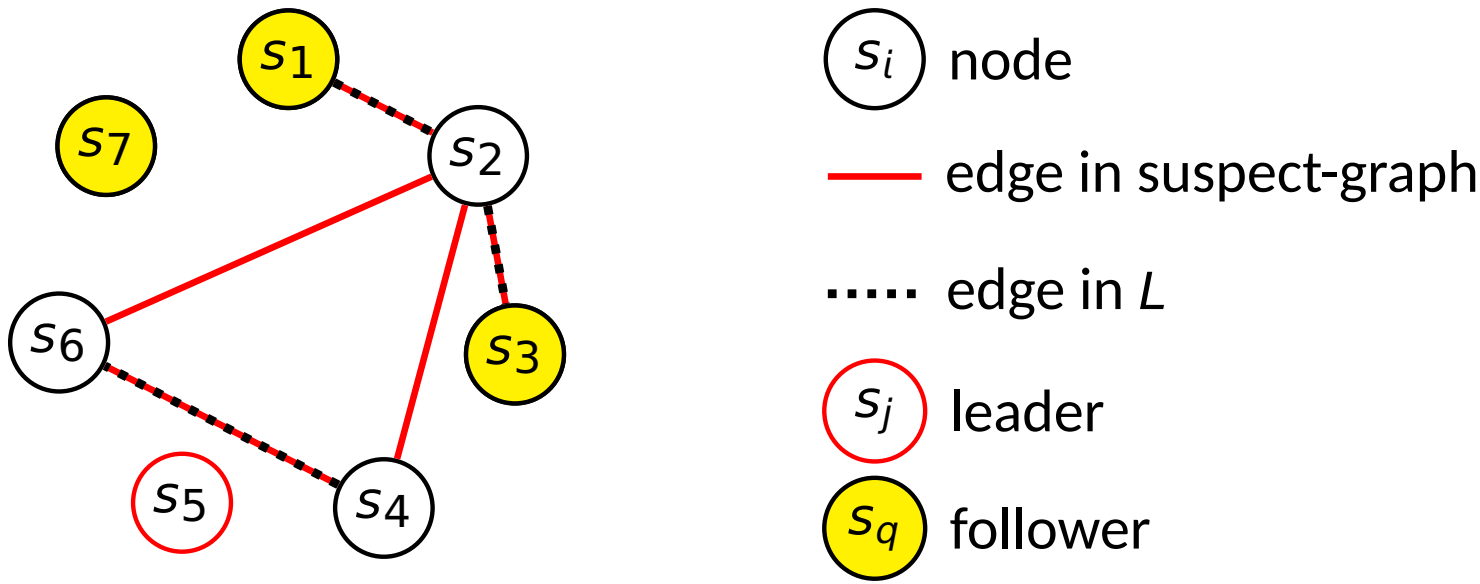


Figure: example graph with $f = 2$

Follower-Selection

- find subgraph L, acyclic with maximum degree 2
- select a leader
  - node with degree 0 in L
- leader selects followers of degree 0 or 1
  - use failure detector to suspect misbehaving leader

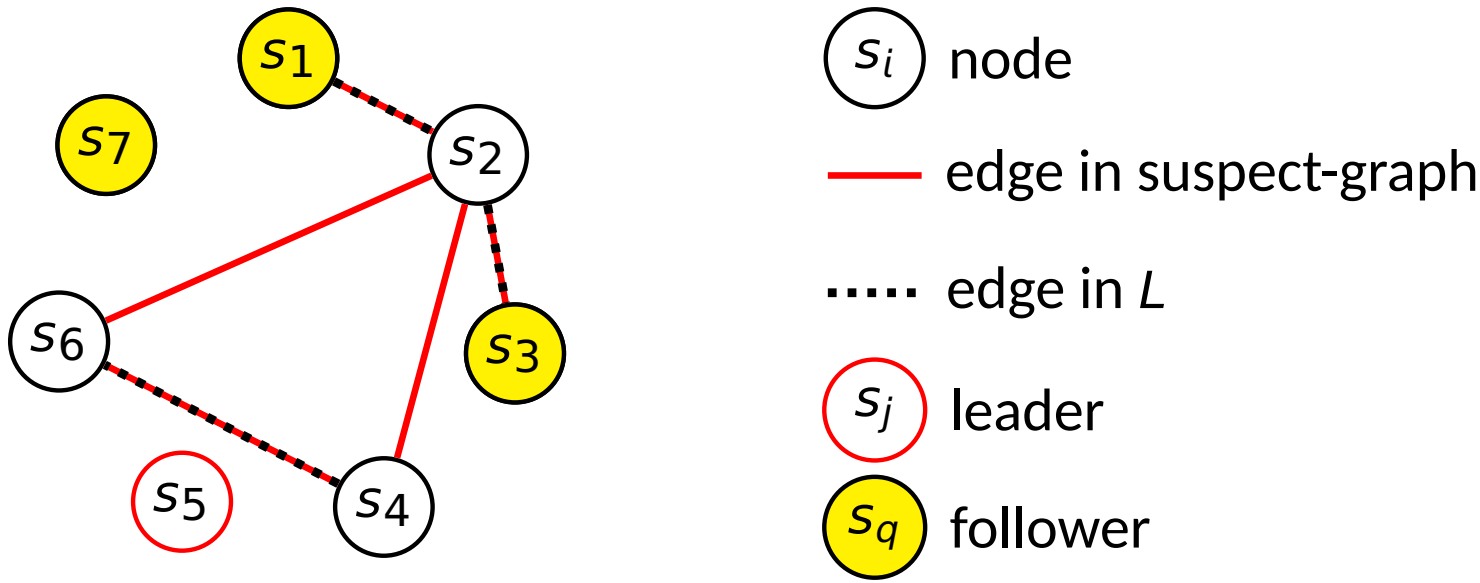


Figure: example graph with $f = 2$

Follower-Selection

- find subgraph L, acyclic with maximum degree 2
- select a leader
  - node with degree 0 in L
- leader selects followers of degree 0 or 1
  - use failure detector to suspect misbehaving leader

at most $6f$ quorums with accurate failure detector