RAFT: FAIL-STOP-RESTART

Attendance

SAFETY, LIVENESS (AVAILABILITY)

Nothing bad happens

Something good happens often

client ⇄ REQ / REPLIES ⇄ D.S

Log

index	1	2	3	
Entry	inst1	inst2	inst3	...
Term			⋮	

$\geq \frac{N}{2}$ followers ACK

See last class for details / heartbeats

Considered

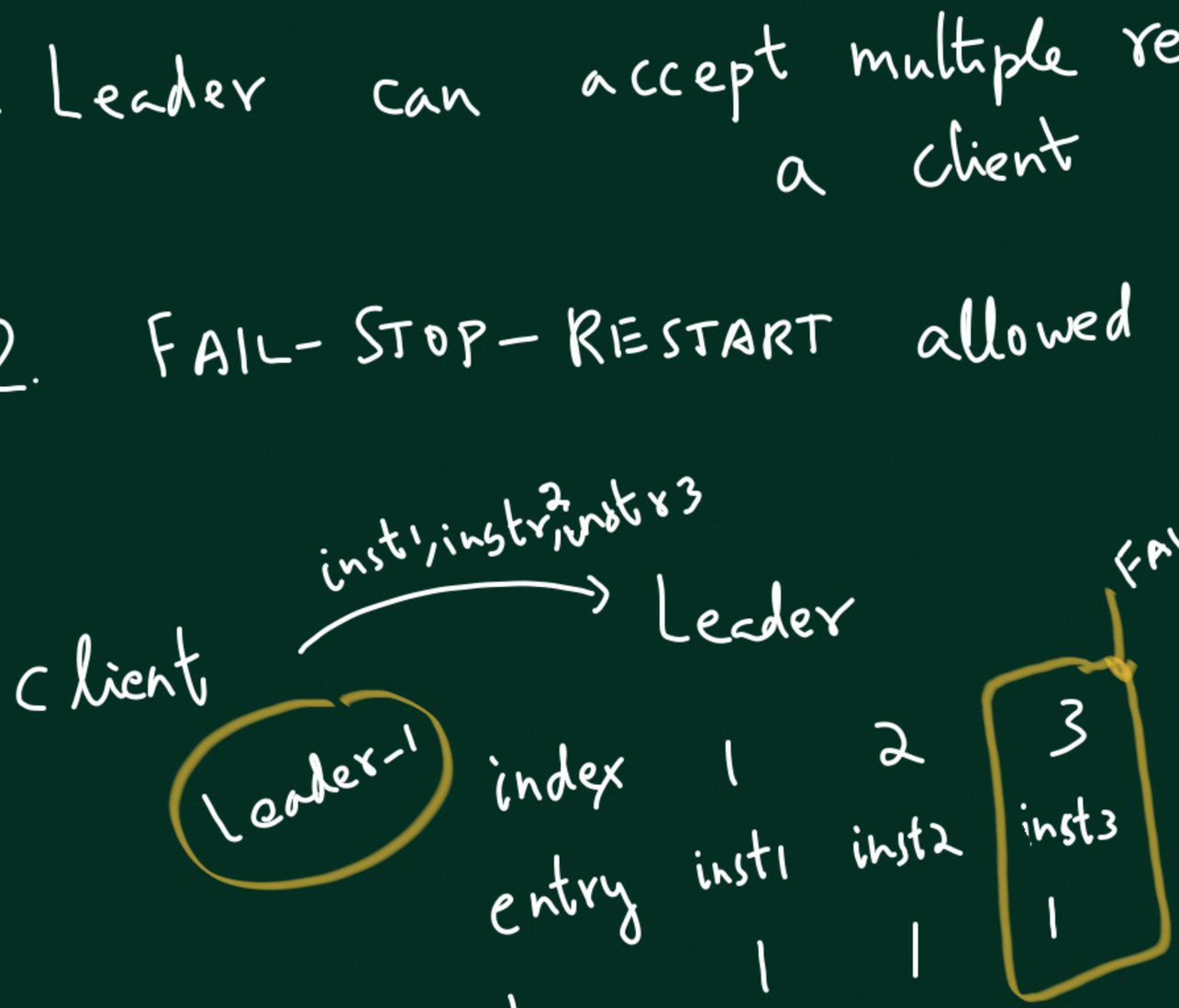
1. Leader only accepts 1 req at a time
2. FAIL-STOP (no restart)

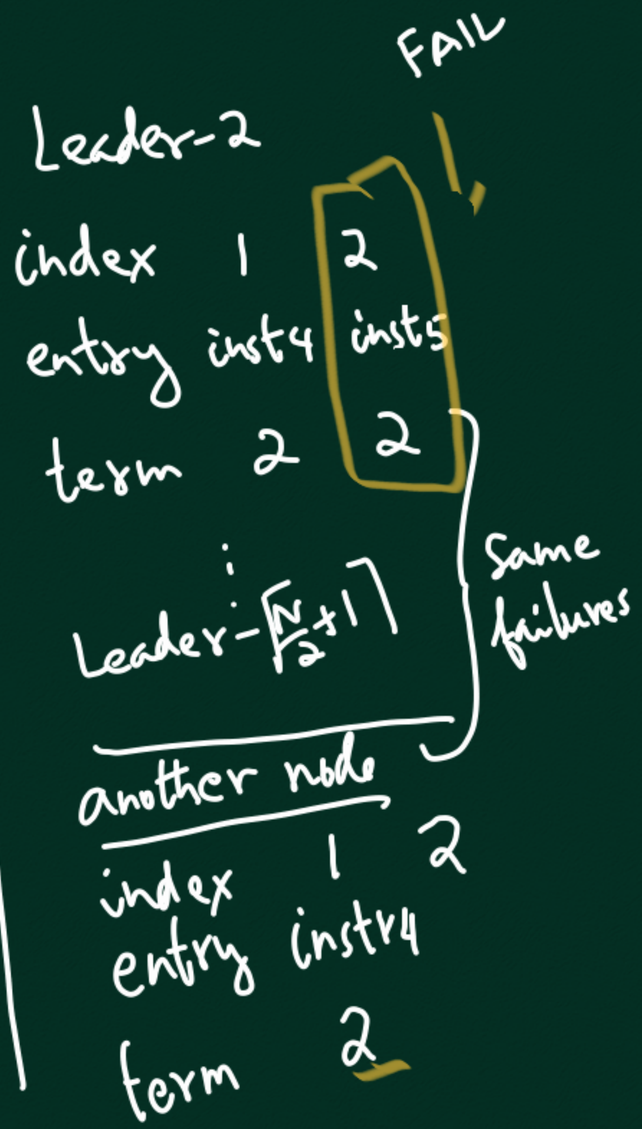
Vote for candidate: if my log entries are a prefix (or exactly same) as a candidate

More General Case

1. Leader can accept multiple requests from a client

2. Fail-Stop-Restart allowed





Note: According to old "prefix rule" the old leaders won't vote for each other as their log entries are very different

	OLD		
index	1	2	3
entry	inst1		

↑ > N/2 ACK

Commit (exec.) and reply

New Rule For Electing Leader

Vote for a candidate if it is more (or as) up-to-date as myself

Rule For Being More Up-To-Date

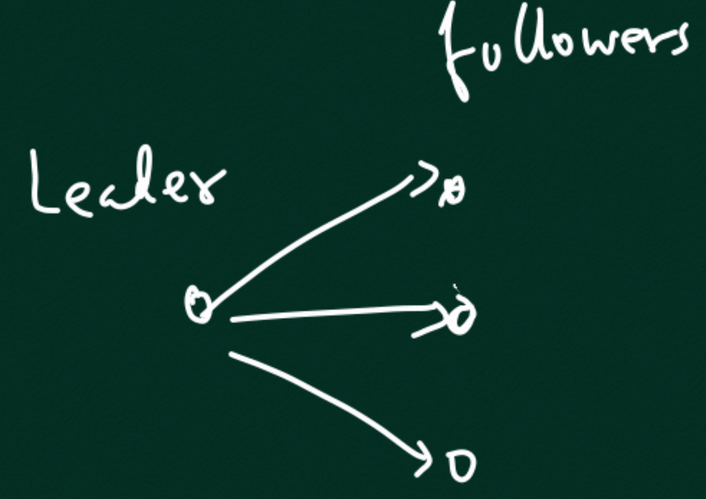
Compare last entries in logs

- if Candidate has a higher term, it is more up-to-date
- if terms are the same, the node with longer log is more up-to-date

Liveness

Broadcast time << (Avg.) Election Timeout

<< Avg. Time Between node failures



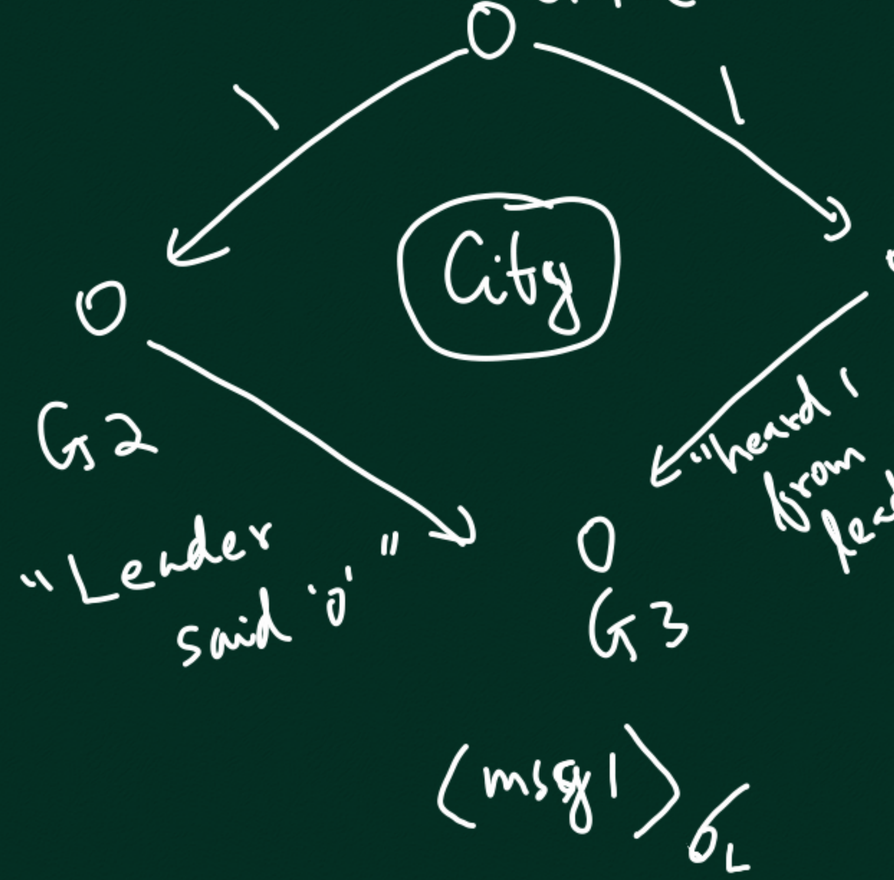
Broadcast Time: Time for leader to send msg to all.

Then we have liveness

Byzantine Failures → any deviation from protocol

1 ⇒ attack
0 ⇒ don't attack

Generals



Want

(1) All honest generals make the same decision

(2) If leader is honest, all honest generals follow his command.

How to solve?

1) Use digital signatures. Assume forging is not possible

↳ replay attack (send old msgs) still possible

2) Use Quorums

(Practical Byz. Fault Tolerance → Liskov, ...) → 1999

Suppose $N = 3f + 1$

then we assume <u>at most</u> f nodes are Byzantine

and the rest are <u>honest</u>

We say we have a Quorum if $2f+1$ nodes (or more) agree on something.

Intersection Property: If set of node Q_1 forms a Quorum, and set Q_2 forms another Quorum then $Q_1 \cap Q_2$ must contain an honest node

Proof: Assume there is no honest node in $Q_1 \cap Q_2$

Hence $|Q_1 \cap Q_2| \leq f \Rightarrow -|Q_1 \cap Q_2| \geq -f$

$$|Q_1 \cup Q_2| = |Q_1| + |Q_2| - |Q_1 \cap Q_2| \geq 2f+1 + 2f+1 - f$$
$$= 3f+2 > N$$

Not possible

Availability Property: Even if all Byz. nodes remain silent the honest nodes are enough in number to potentially form a Quorum

SAFETY → Same as RAFT
→ guaranteed even if network is Asynchronous

LIVENESS: Clients eventually receive replies to their requests

achieved when? If a msg is sent out for for the first time at time t, then it is received by others by $t + o(t)$ → delay

Impossibility Result: FLP, CAP, ...

(1985) FISHER, LYNCH, PATTERSON: It is impossible to have a deterministic protocol that solves consensus in a message passing asynchronous system in which at most one process may fail by crashing (if they reach a decision)

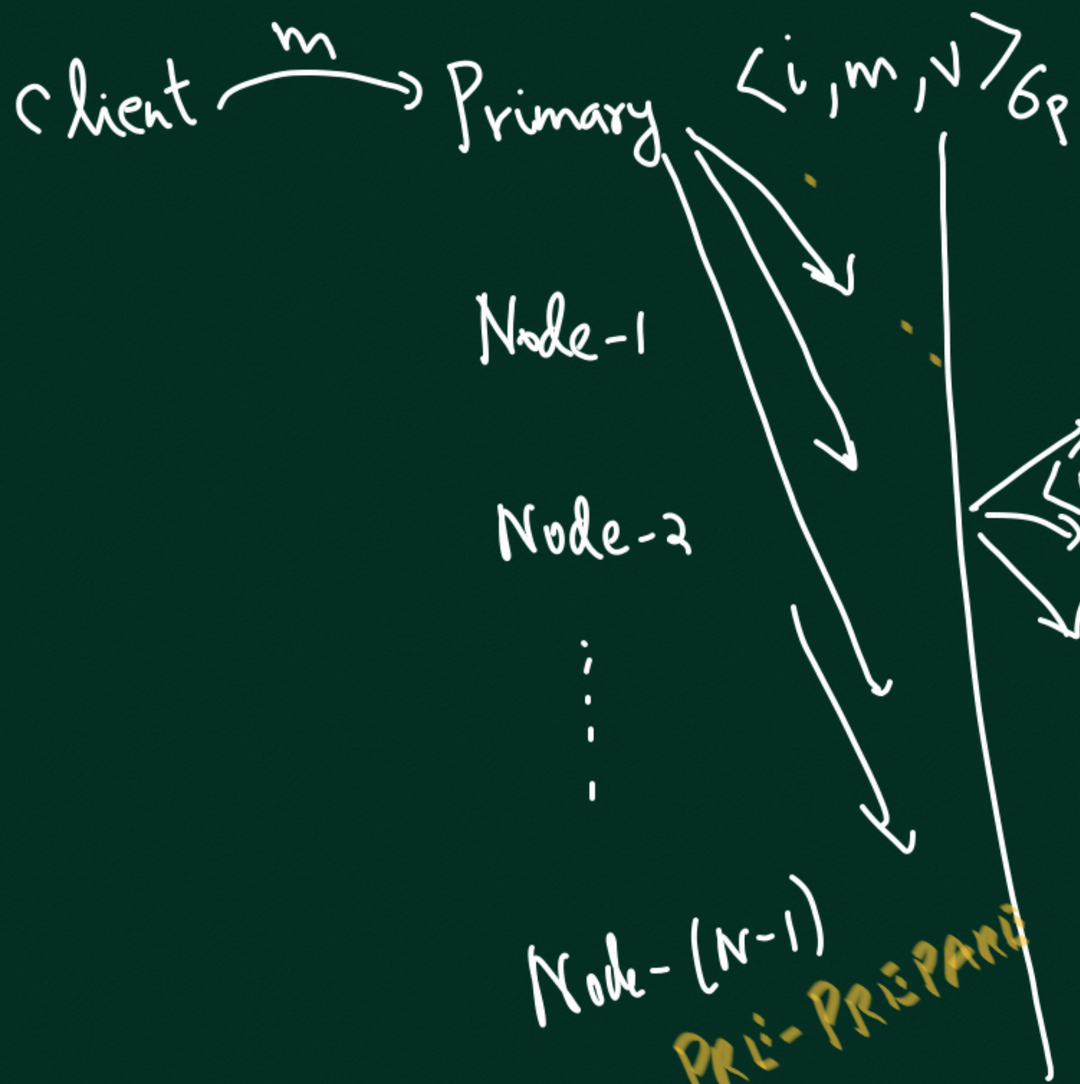
STRONG CONSENSUS

a) (AGREEMENT) All honest gens. must reach same decision

(b) (Termination) All honest gens. must eventually make a decision

(c) (Validity) If Gen1 (leader) is honest, then all honest gens. must agree to his command

High Level idea of PBFT



Log

Leader → Primary

(Raft) (PBFT)

index (i)

entry (message, m)

term (VIEW, v)

i

m

v