CLOUD COMPUTING CONCEPTS

with Indranil Gupta (Indy)

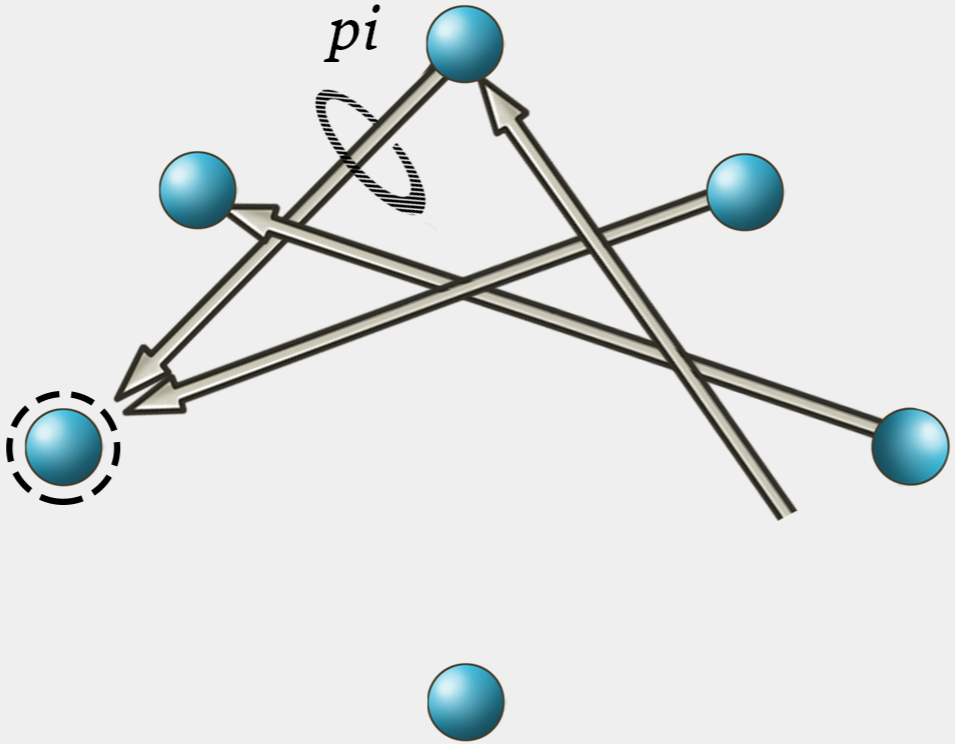
MEMBERSHIP

Lecture C

GOSSIP-STYLE MEMBERSHIP

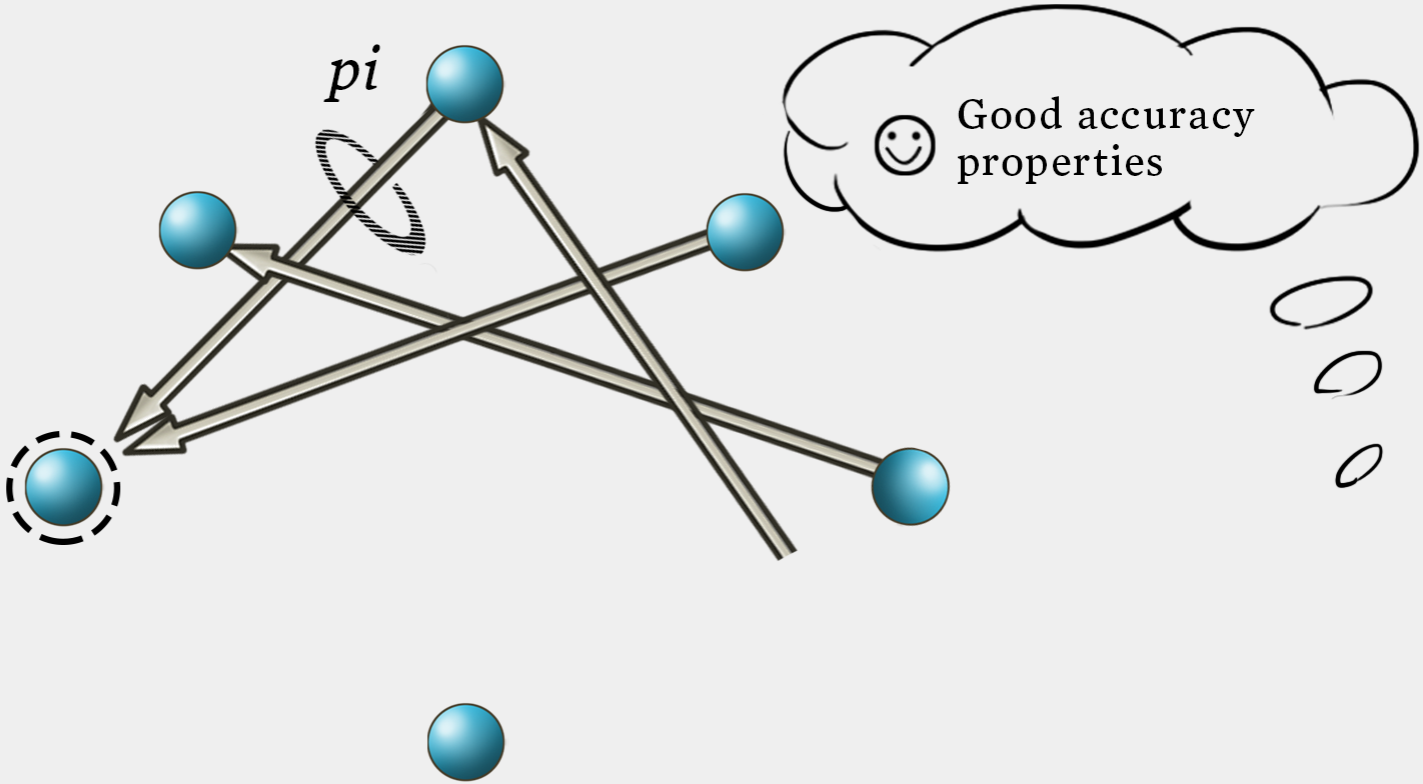
GOSSIP-STYLE HEARTBEATING

Array of
Heartbeat seq. /
for member subset

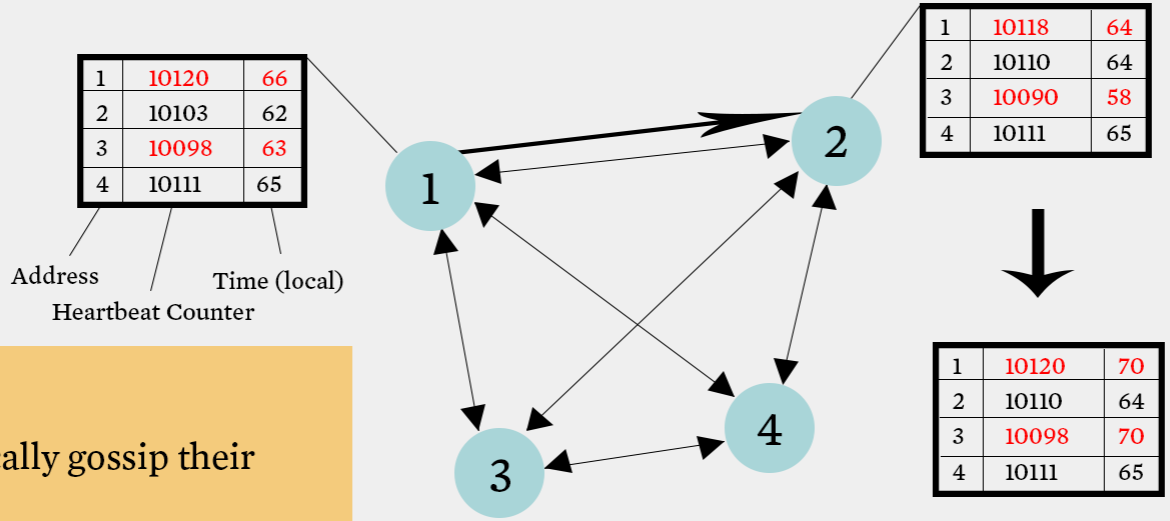


GOSSIP-STYLE HEARTBEATING

Array of
Heartbeat seq. /
for member subset



GOSSIP-STYLE FAILURE DETECTION



Protocol

- Nodes periodically gossip their membership list
- On receipt, the local membership list is updated

Current time: 70 at node 2 (asynchronous clocks)

Fig and animation by: Dongyun Jin and Thuy Ngyuen

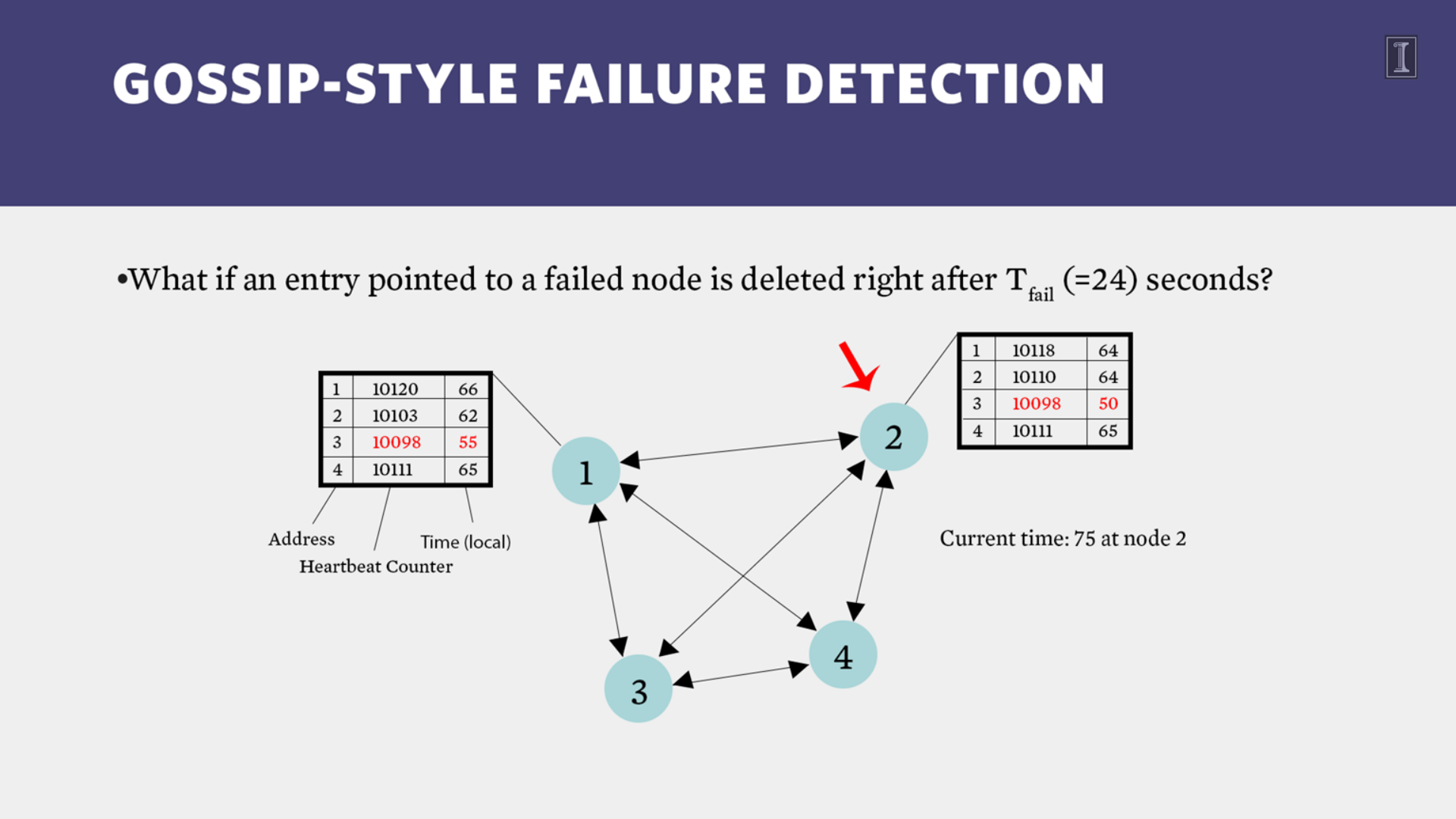
GOSSIP-STYLE FAILURE DETECTION

- If the heartbeat has not increased for more than T_{fail} seconds, the member is considered failed

- And after $T_{cleanup}$ seconds, it will delete the member from the list

- Why two different timeouts?

GOSSIP-STYLE FAILURE DETECTION



- What if an entry pointed to a failed node is deleted right after T_{fail} (=24) seconds?

1	10120	66
2	10103	62
3	10098	55
4	10111	65

Address

Heartbeat Counter

Time (local)

1

2

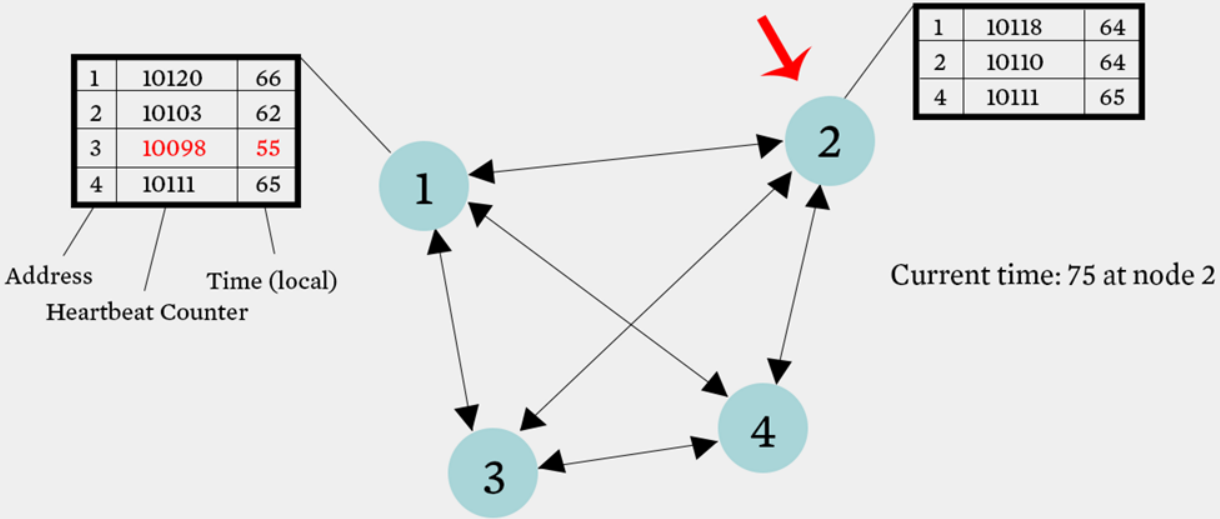
3

4

1	10118	64
2	10110	64
3	10098	50
4	10111	65

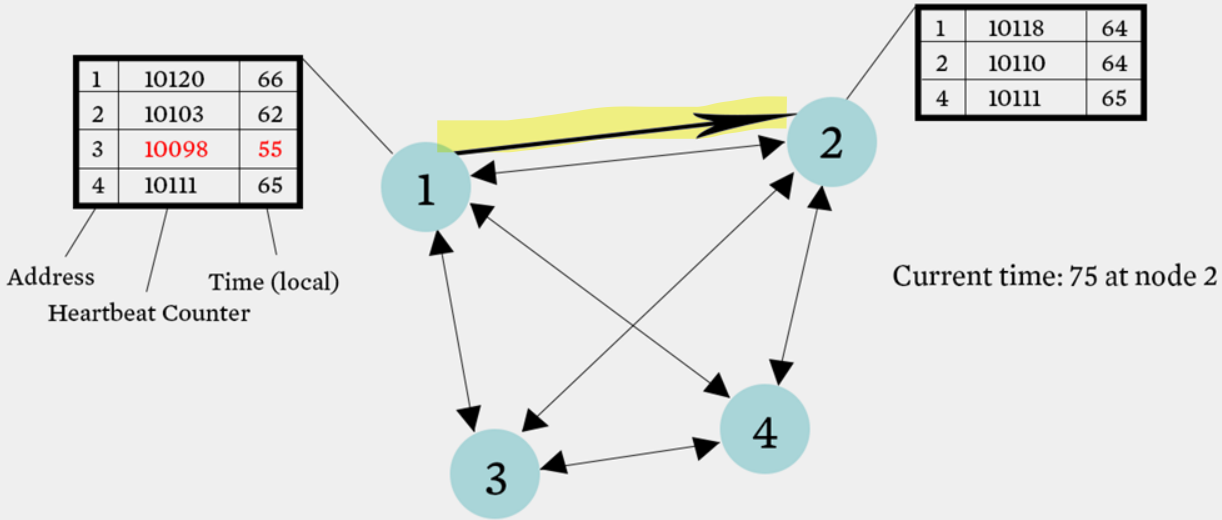
Current time: 75 at node 2

GOSSIP-STYLE FAILURE DETECTION



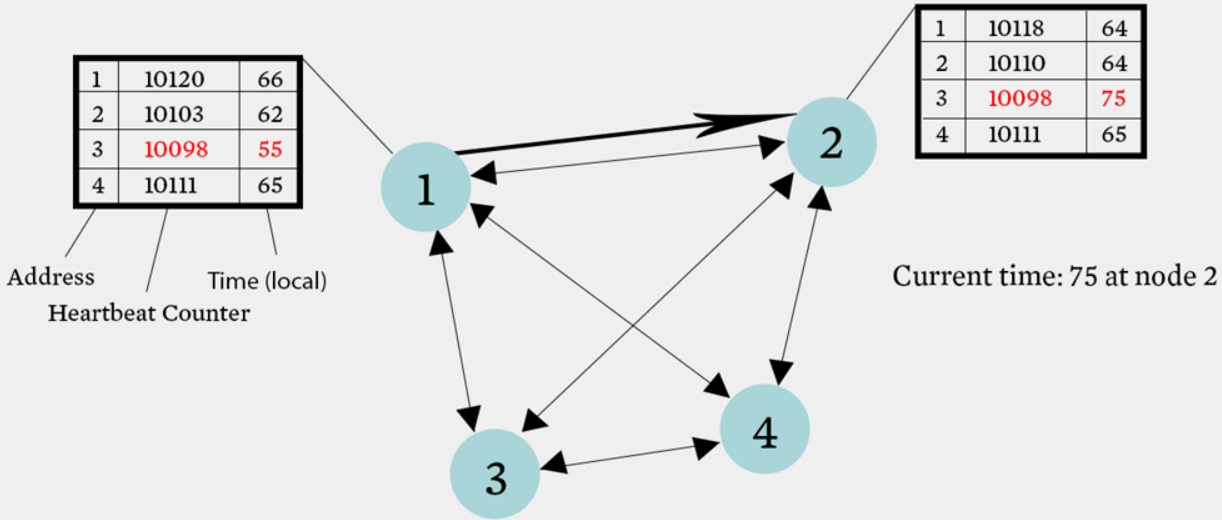
- What if an entry pointed to a failed node is deleted right after T_{fail} (=24) seconds?

GOSSIP-STYLE FAILURE DETECTION

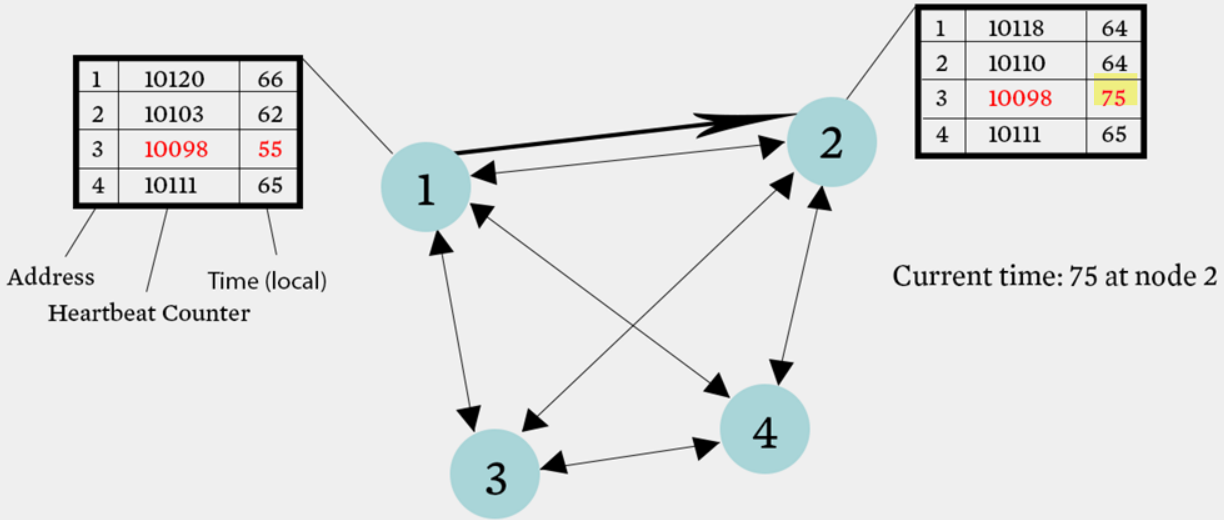


- What if an entry pointed to a failed node is deleted right after T_{fail} (=24) seconds?

GOSSIP-STYLE FAILURE DETECTION



GOSSIP-STYLE FAILURE DETECTION



- What if an entry pointed to a failed node is deleted right after T_{fail} (=24) seconds?

- Fix: remember for another T_{fail}

MULTI-LEVEL GOSSIPING

- Network topology is hierarchical
- Random gossip target selection => core routers face O(N) load (Why?)

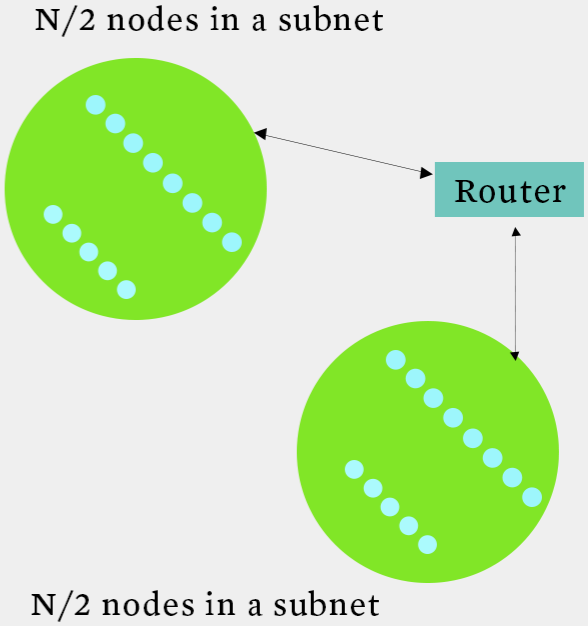
- Fix: Select gossip target in subnet *i*, which contains n_i nodes, with probability $1/n_i$

- Router load=O(1)

- Dissemination time=O(log(N))
  - Why?

- What about latency for multi-level topologies?

[Gupta et al, TPDS 06]



ANALYSIS/DISCUSSION

- What happens if gossip period T_{gossip} is decreased?
- A single heartbeat takes O(log(N)) time to propagate.

  So: N heartbeats take:
  - O(log(N)) time to propagate, if bandwidth allowed per node is allowed to be O(N)
  - O(N.log(N)) time to propagate, if bandwidth allowed per node is only O(1)
  - What about O(k) bandwidth?
- What happens to $P_{mistake}$ (false positive rate) as T_{fail} ,$T_{cleanup}$ is increased?
- Tradeoff: False positive rate vs. detection time vs. bandwidth

NEXT

- So, is this the best we can do? What is the best we can do?