CLOUD COMPUTING CONCEPTS

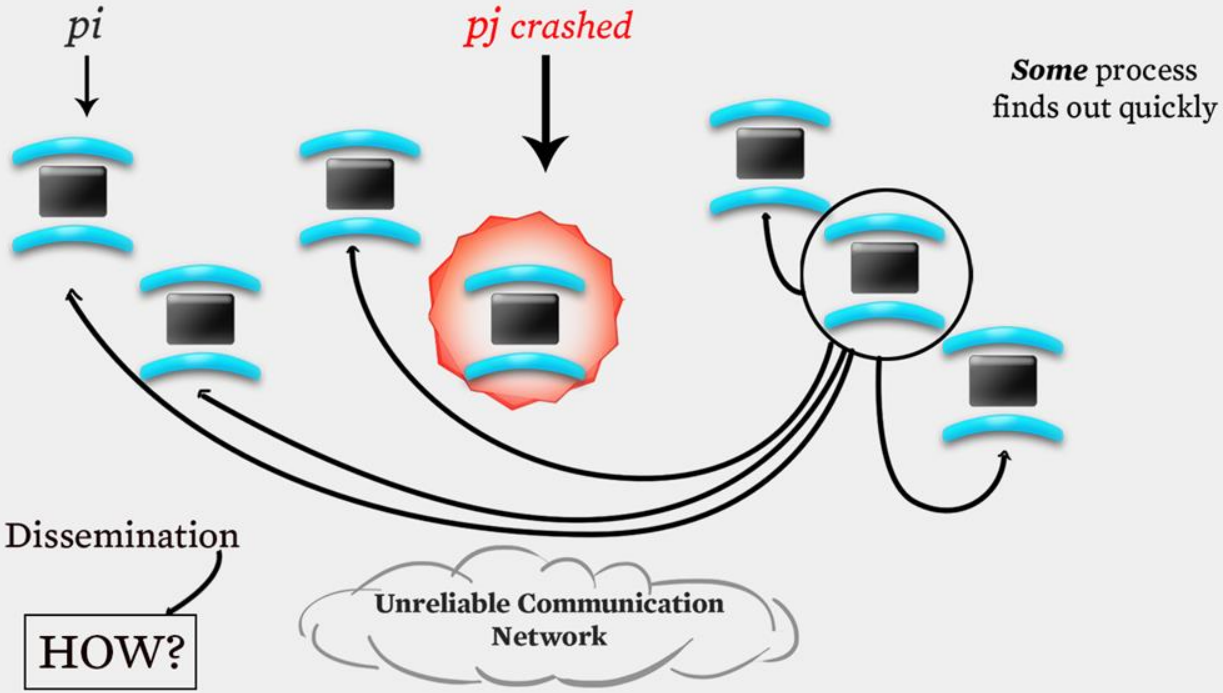
with Indranil Gupta (Indy)

MEMBERSHIP

Lecture F

DISSEMINATION AND SUSPICION

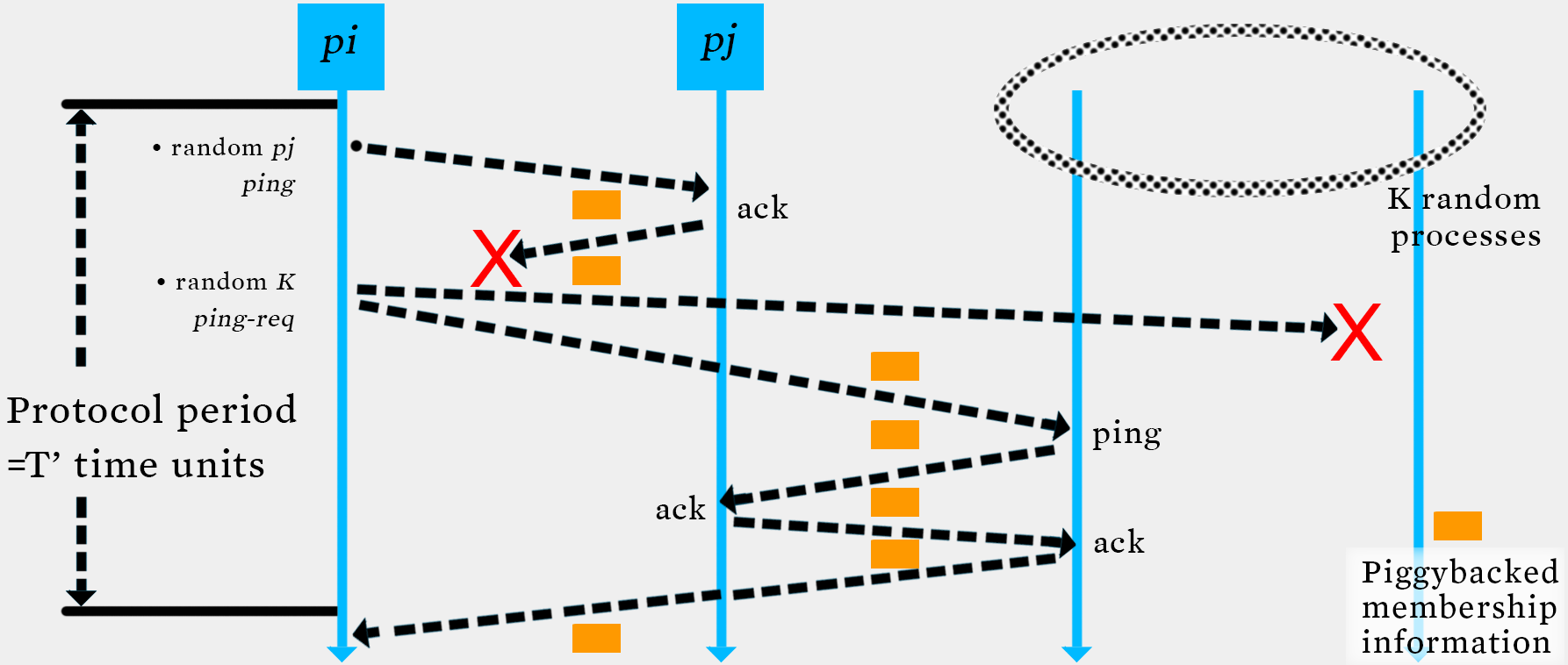
III. DISSEMINATION



DISSEMINATION OPTIONS

- Multicast (Hardware / IP)
  - unreliable
  - multiple simultaneous multicasts
- Point-to-point (TCP / UDP)
  - expensive
- Zero extra messages: Piggyback on Failure Detector messages
  - Infection-style Dissemination

SWIM FAILURE DETECTOR PROTOCOL



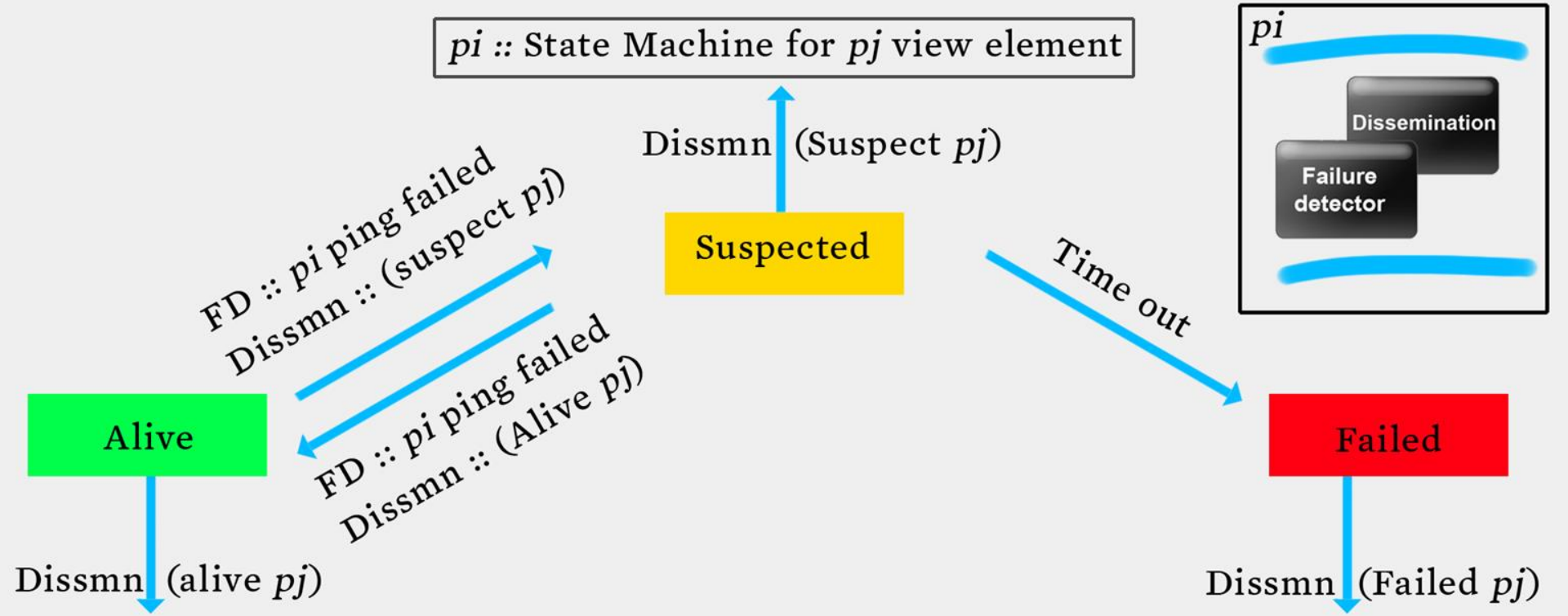
INFECTION-STYLE DISSEMINATION

- Epidemic style dissemination
  - After $\lambda.\log(N)$ protocol periods, $N^{-(2\lambda-2)}$ processes would not have heard about an update
- Maintain a buffer of recently joined/evicted processes
  - Piggyback from this buffer
  - Prefer recent updates
- Buffer elements are garbage collected after a while
  - After $\lambda.\log(N)$ protocol periods; this defines weak consistency

SUSPICION MECHANISM

- False detections, due to:
  - Perturbed processes
  - Packet losses, e.g., from congestion
- Indirect pinging may not solve the problem
  - e.g., correlated message losses near pinged host
- Key: *suspect* a process before *declaring* it as failed in the group

SUSPICION MECHANISM



SUSPICION MECHANISM

- Distinguish multiple suspicions of a process
  - Per-process *incarnation number*
  - *Inc #* for pi can be incremented only by pi
    - e.g., when it receives a (Suspect, pi) message
  - Somewhat similar to DSDV

- Higher inc# notifications over-ride lower inc#'s
- Within an inc#: (Suspect inc #) > (Alive, inc #)
- (Failed, inc #) overrides everything else

WRAP UP

- Failures the norm, not the exception in datacenters
- Every distributed system uses a failure detector
- Many distributed systems use a membership service

- Ring failure detection underlies
  - IBM SP2 and many other similar clusters/machines

- Gossip-style failure detection underlies
  - Amazon EC2/S3 (rumored!)