



CLOUD COMPUTING CONCEPTS

with Indranil Gupta (Indy)

MEMBERSHIP

Lecture A

WHAT IS GROUP MEMBERSHIP LIST?

A CHALLENGE

- You've been put in charge of a datacenter, and your manager has told you, "Oh no! We don't have any failures in our datacenter!"
- Do you believe him/her?
- What would be your first responsibility?
- Build a failure detector
- What are some things that could go wrong if you didn't do this?

FAILURES ARE THE NORM

... not the exception, in datacenters.

- Say, the rate of failure of one machine (OS/disk/motherboard/network, etc.) is once every 10 years (120 months) on average.
- When you have 120 servers in the DC, the **mean time to failure (MTTF)** of the next machine is 1 month.
- When you have 12,000 servers in the DC, the MTTF is about once every 7.2 hours!

A CHALLENGE

- You've been put in charge of a datacenter, and your manager has told you, "Oh no! We don't have any failures in our datacenter!"
- Do you believe him/her?
- What would be your first responsibility?
- Build a failure detector
- What are some things that could go wrong if you didn't do this?

TO BUILD A FAILURE DETECTOR

- You have a few options:
 1. Hire 1000 people, each to monitor one machine in the datacenter and report to you when it fails.
 2. Write a failure detector program (distributed) that automatically detects failures and reports to your workstation.

Which is more preferable, and why?

TARGET SETTINGS

- Process ‘group’-based systems
 - Clouds/Datacenters
 - Replicated servers
 - Distributed databases
- Crash-stop/Fail-stop process failures

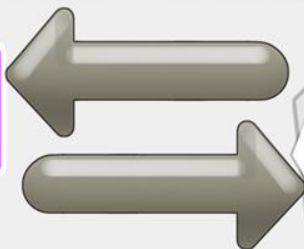
GROUP MEMBERSHIP SERVICE

Application Queries
e.g., gossip. overlays,
DHT's, etc.

Application Process p_i



Membership List



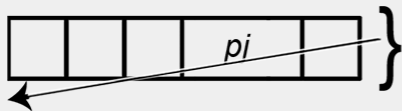
Membership
Protocol



Unreliable
Communication

TWO SUB-PROTOCOLS

Application Process p_i



- **Complete list all the time (strongly consistent)**
 - Virtual synchrony
- **Almost-Complete list (weakly consistent)**
 - Gossip-style, SWIM, ... (FOCUS OF THIS LECTURE SERIES)
- **Or Partial-random list (other systems)**
 - SCAMP, T-MAN, Cyclon, ...

Dissemination

Failure
detector

Unreliable
Communication

LARGE GROUP: SCALABILITY A GOAL

this is us (pi)



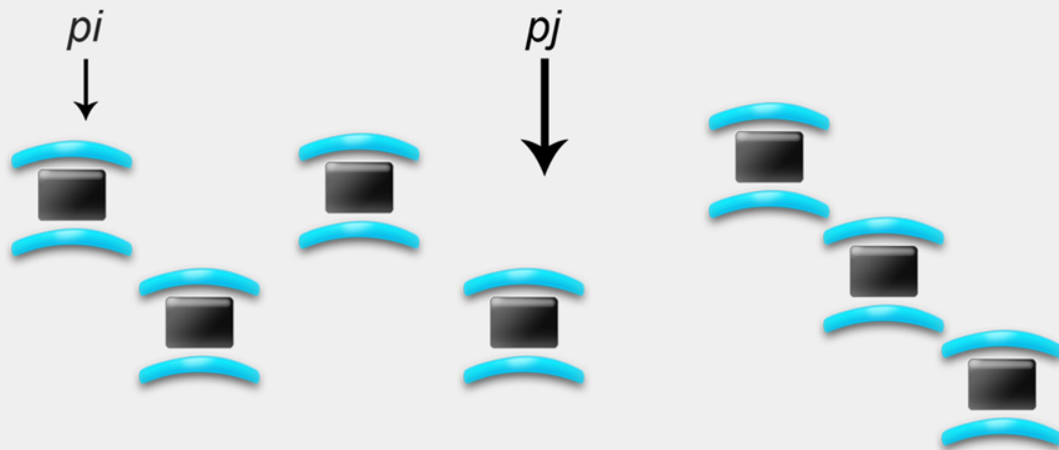
Process Group
"Members"

1000's of processes

Unreliable Communication
Network



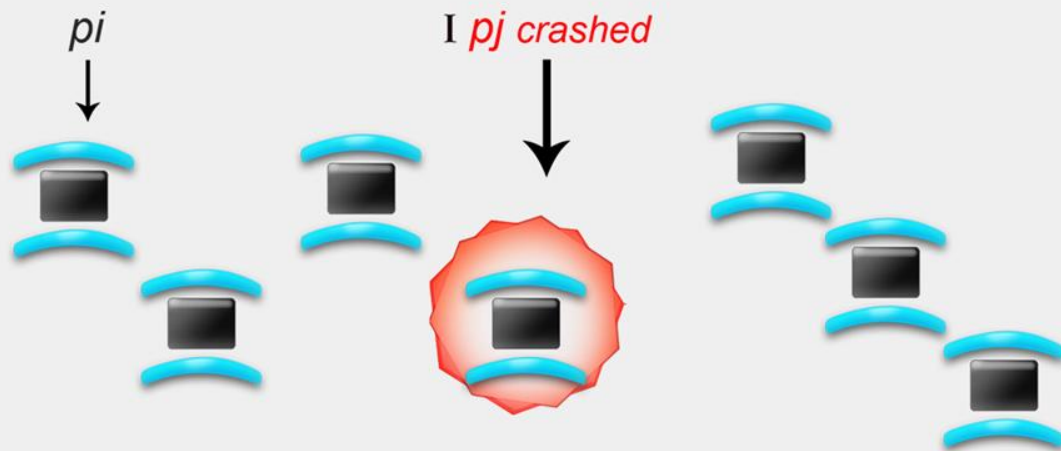
GROUP MEMBERSHIP PROTOCOL



Crash-stop Failures only

Unreliable Communication
Network

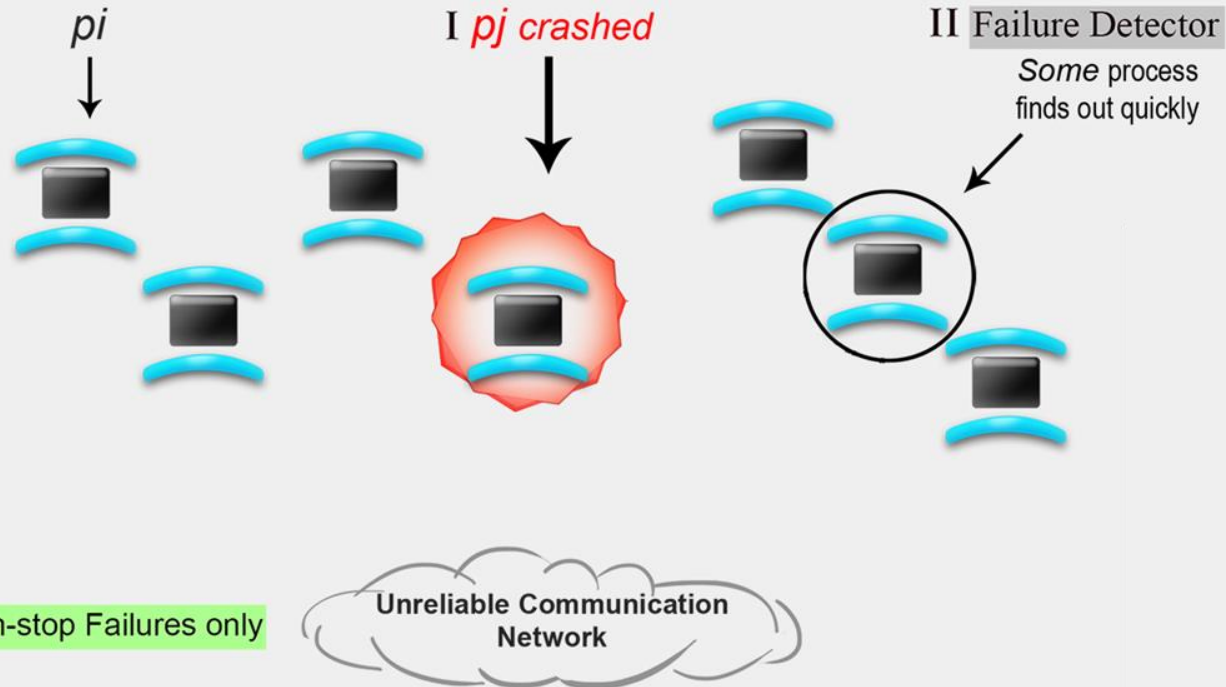
GROUP MEMBERSHIP PROTOCOL



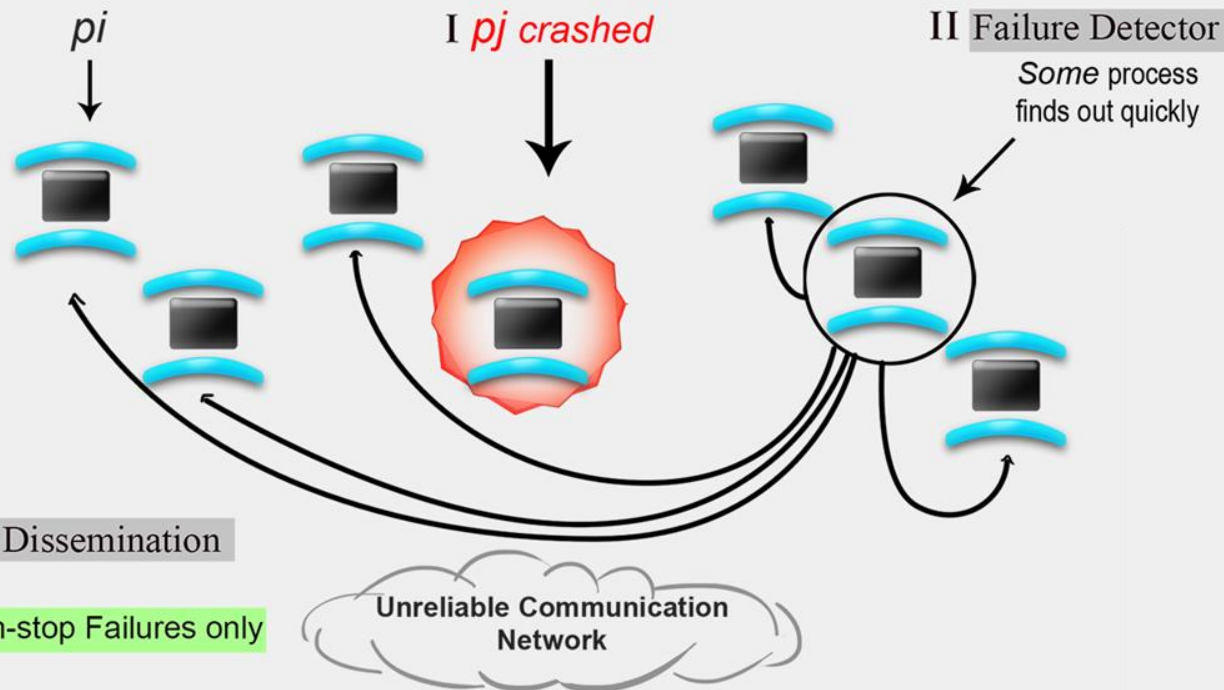
Crash-stop Failures only

Unreliable Communication
Network

GROUP MEMBERSHIP PROTOCOL



GROUP MEMBERSHIP PROTOCOL



NEXT

- How do you design a group membership protocol?