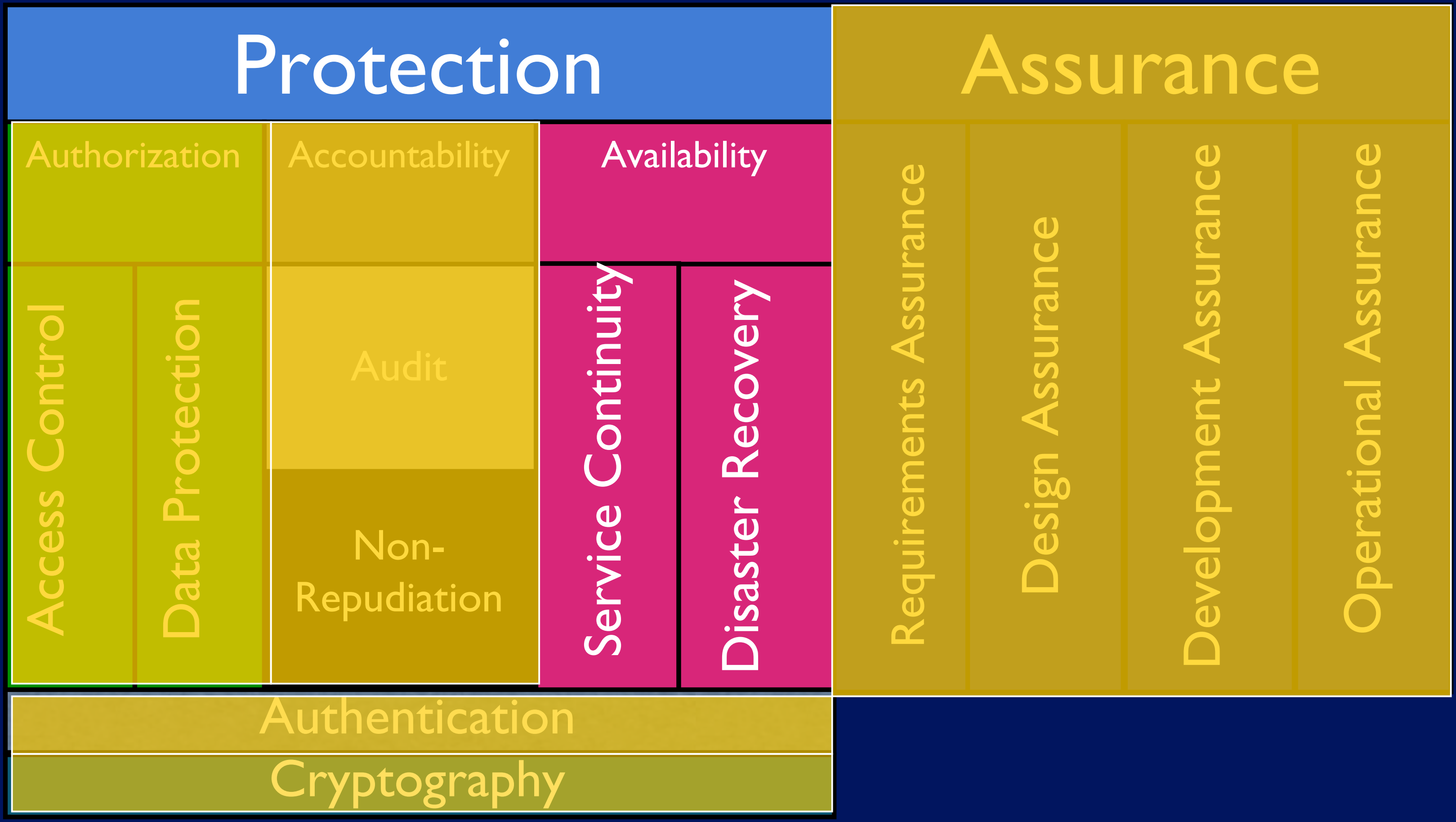
THE UNIVERSITY OF BRITISH COLUMBIA

# Availability

EECE 412

# Where We Are



Protection

Authorization

Access Control

Data Protection

Accountability

Audit

Non-Repudiation

Availability

Service Continuity

Disaster Recovery

Authentication

Cryptography

Assurance

Requirements Assurance

Design Assurance

Development Assurance

Operational Assurance

# What do you already know?

- How are **error**, **fault**, and **failure** different?
- What’s the difference between **fail-stop** and **Byzantine** failures?
- How many nodes do you need to have **3-fault** tolerance for **Byzantine** failures?
- What measures to deal with failures do you know?
- What are the ways of achieving service continuity in the presence of attacks?

# Outline

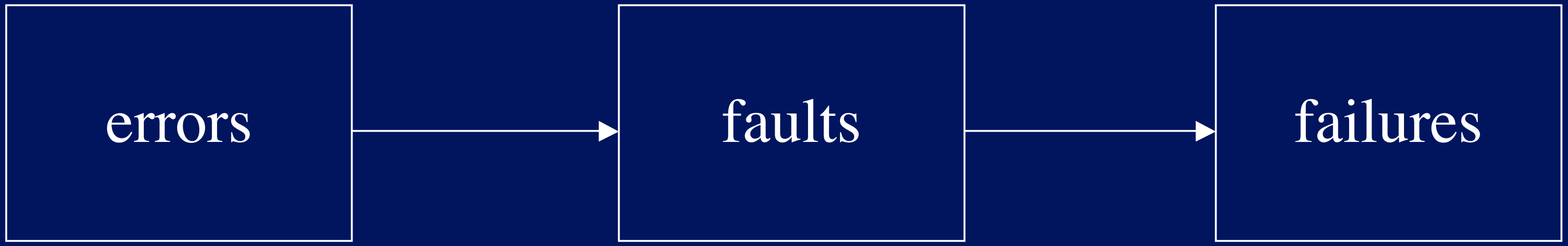
- Availability in the presence of failures
  - FT terminology
  - k fault tolerance
  - two army problem
  - Byzantine Generals problem
  - Services continuity and disaster recovery
- Availability in the presence of attacks
  - Failures vs. attacks
  - Random vs. scale-free networks
  - Internet tolerance to attacks and failures
  - Services continuity and disaster recovery

THE UNIVERSITY OF BRITISH COLUMBIA

# Availability in the Presence of Failures

# Failures, Errors, and Faults

- A system is said to fail when it cannot meet its promises
- Error may lead to a fault
- Fault -- a cause of an error


errors

faults

failures

# Fault Types

- Transient: occur once and then disappear

- Intermittent: occurs, then vanishes, then reappears

- Permanent: continues to exist

# Availability and Reliability

- **Availability**: Probability that a system operates correctly at any given moment and is available to perform its functions

- **Reliability**: time period during which a system continues to be available to perform its functions

- Problem: calculate system availability and reliability if it’s unavailable for 1 second every hour.

# Fault Tolerance

A fault tolerant system can provide its services even in the presence of faults

# Classification of Failure Modes

| Type of failure | Description |
|---|---|
| Crash failure | A server halts, but is working correctly until it halts |
| Omission failure | A server fails to respond to incoming requests |
| Receive omission | A server fails to receive incoming messages |
| Send omission | A server fails to send messages |
| Timing failure | A server's response lies outside the specified time interval |
| Response failure | The server's response is incorrect |
| Value failure | The value of the response is wrong |
| State transition failure | The server deviates from the correct flow of control |
| Arbitrary (a.k.a. Byzantine) failure | A server may produce arbitrary responses at arbitrary times |

# Achieving k fault tolerance

A system is k fault tolerant if it can survive faults in k components

- silent failure vs. Byzantine failure

  k+1 2k+1

# Agreement among honest players with unreliable communications: Two-army Problem

Even with nonfaulty processes, agreement even between two processes is not possible in the face of unreliable communications

# Agreement among dishonest players with perfect communications: Byzantine Generals Problem

Results:

1. In a system with *m* faulty processes, agreement can be achieved only if *2m+1* correctly functioning processes are present (total 3m+1). (Lamport et al., 1982)

2. If messages cannot be guaranteed to be delivered within a known, finite time, no agreement is possible even with one faulty process. (Fischer et al., 1985)

# Ways to Deal with Failures

- Service continuity
  - Masking failures via
    - Redundancy of
      - information
      - time
      - physical
- Disaster recovery
  - Backward recovery
    - check pointing
  - Forward recovery
    - bringing system into a correct new state
  - Don’t underestimate backups!

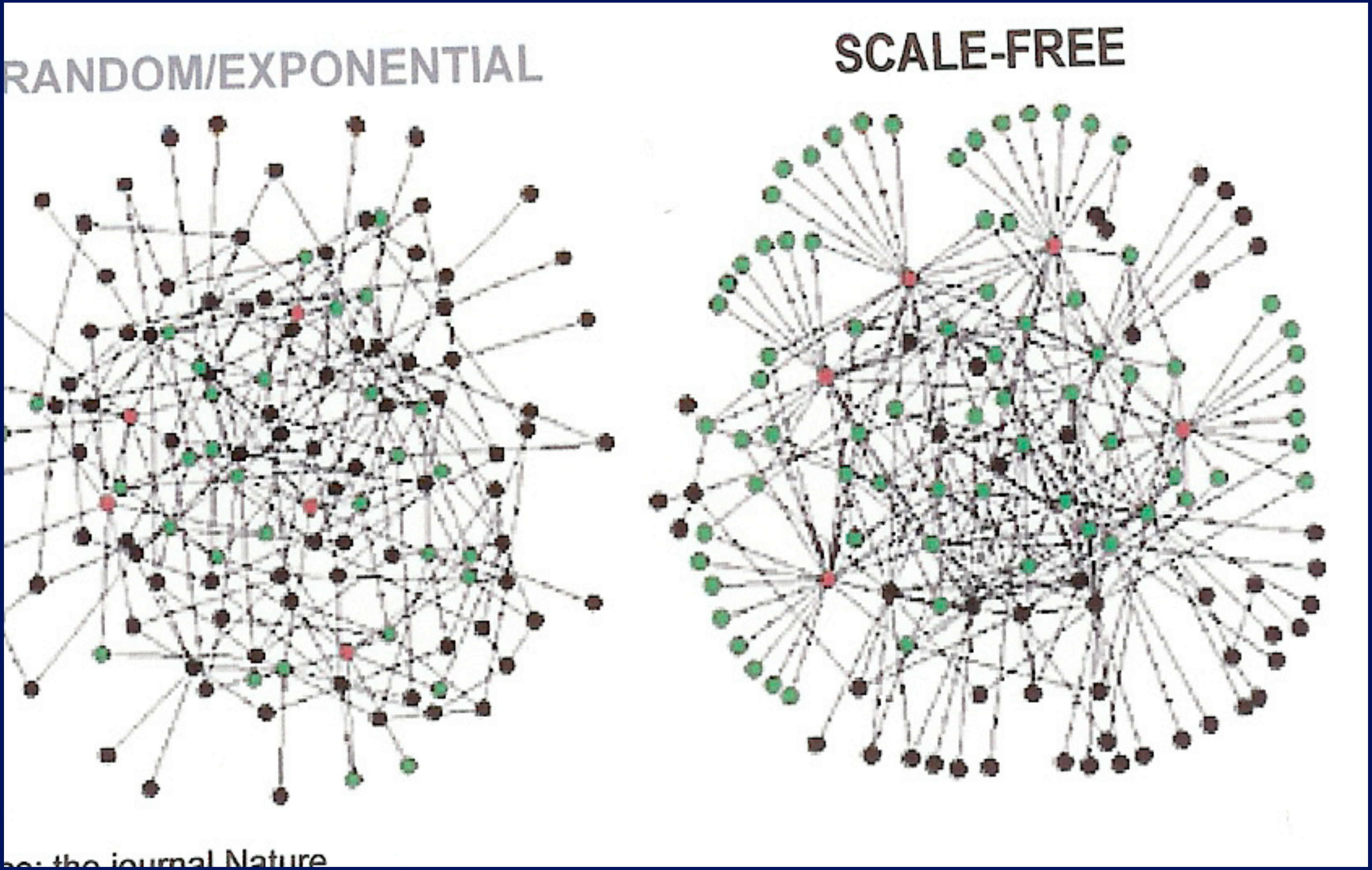
THE UNIVERSITY OF BRITISH COLUMBIA

# Availability in the Presence of Attacks

# Failures vs. Attacks

- **Failure**
  - Random (unintentional) unavailability of participants and/or infrastructure elements
- **Attack**
  - Systematic (intentional) unavailability of participants and/or infrastructure elements

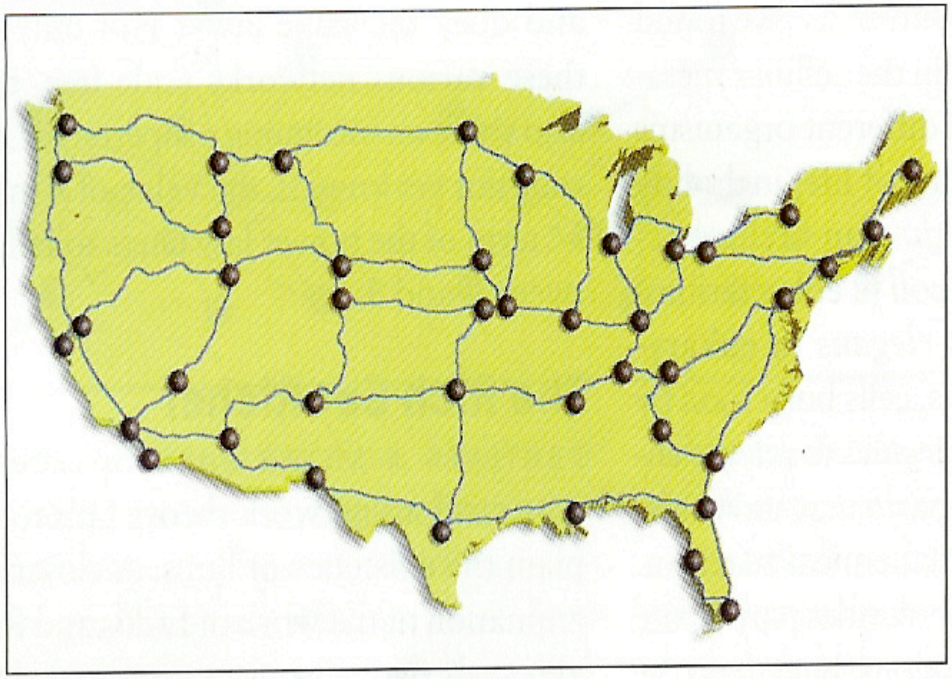
# Random vs. Scale-free Networks



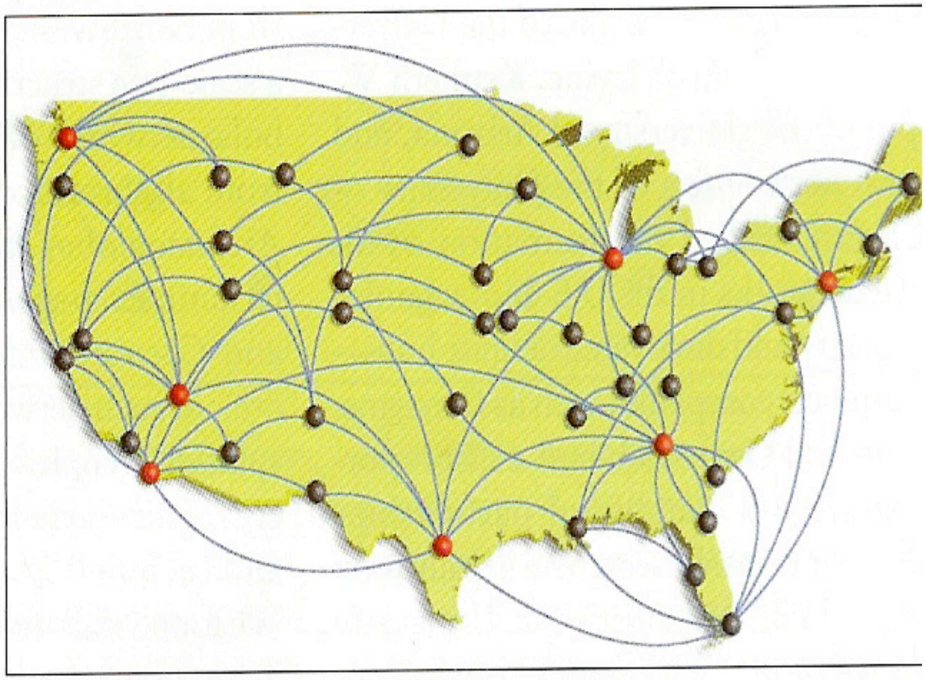
ce: the journal Nature

airline system (simplified in right map),

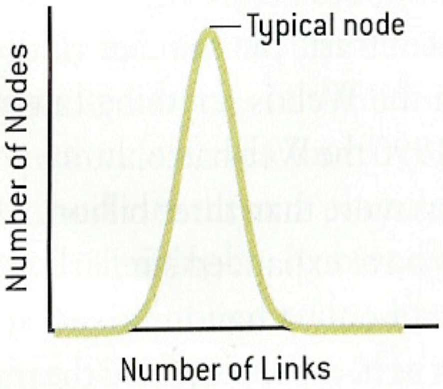
Random Network



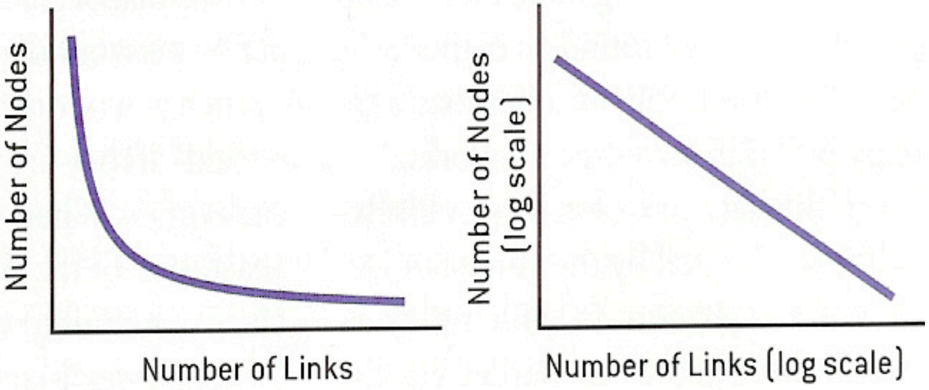
Scale-Free Network



Bell Curve Distribution of Node Linkages

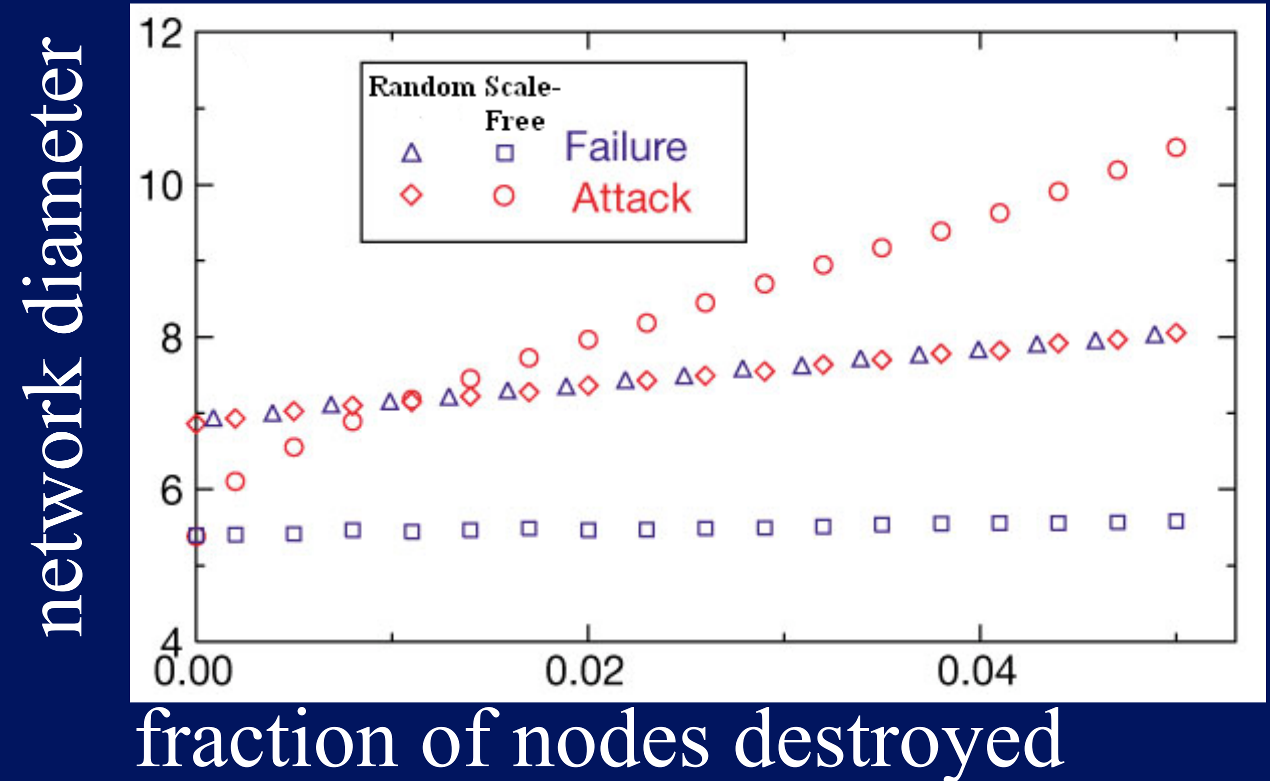


Power Law Distribution of Node Linkages



# Internet Tolerance to Attacks and Failures

- Scale-free networks are failure-tolerant
- Random networks are attack-tolerant



Source: R. Albert, H. Jeong, and A.-L. Barabasi, "Error and attack tolerance of complex networks," Nature, vol. 406, no. 6794, 2000, pp. 378-82.

# Ways to Deal with Attacks

- Service continuity
  - Same as for FT, plus
  - Heterogeneity
    - Diversification
      - Avoid monocultures
    - Randomization
      - Avoid "hubs"
- Disaster recovery
  - Same as for FT

# Summary

- Availability in the presence of failures
  - FT terminology
  - k fault tolerance
  - two army problem
  - Byzantine Generals problem
  - Services continuity and disaster recovery
- Availability in the presence of attacks
  - Failures vs. attacks
  - Random vs. scale-free networks
  - Internet tolerance to attacks and failures
  - Services continuity and disaster recovery

# What did you learn?

- How are **error**, **fault**, and **failure** different?
- What's the difference between **fail-stop** and **Byzantine** failures?
- How many nodes do you need to have **3-fault** tolerance for **Byzantine** failures?
- What measures to deal with failures do you know?
- What are the ways of achieving service continuity in the presence of attacks?