

# Storage Systems

## NPTEL Course

### Jan 2012

(Lecture 02)

## K. Gopinath

## Indian Institute of Science

# Introduction

- Why Storage Systems?
  - Earlier: (processing + storage) and networking
  - Now: processing, storage and networking
  - Fast networks enable separation from “processing”
- Devices, Protocols, Layers/Systems
  - Old and New Devices: Tape, Drum, Disk, Solid State
  - Protocols: NFS, Cloud storage API
  - Layers/Systems: Google FS, Mail storage
- Issues
  - Older: concurrency with CPU, handling device diversity
  - Newer: scale, distribution, error mgmt, security, RT, QoS, manageability

# Why is storage different?

- Consider long term storage (multiple decades): Stored data can be accessed decades later!
  - Formats, devices, etc can change
  - Data not interpretable unless auxiliary information also stored (Recursion problem!)
- Consider security: Why storage security different from, say, network security?
  - Network security is across “space”
    - Network transfers happen within a short time (unless space probe netw packets to Pluto!)
  - Storage security is across both “space and time”
    - If using keys, keys may have to survive years!

# Naming and Storing

- Two important functions:
  - Give a name to an object
    - Can involve some processing (eg. add name to an index)
    - Or, name itself may be computed based on contents
  - Store an object
    - May involve other reads or stores!
    - May involve significant computation
      - Compression, encryption, coding, or deduplication: remove redundancy
- May need to keep aux. information about object
  - Metadata vs data; recursion? (metadata about metadata...)
  - Metadata loss vs data loss
- Access (r/w): device specific aspects determine speed
  - Reads sequential, non-sequential or random
  - Writes in-place or out-of-place