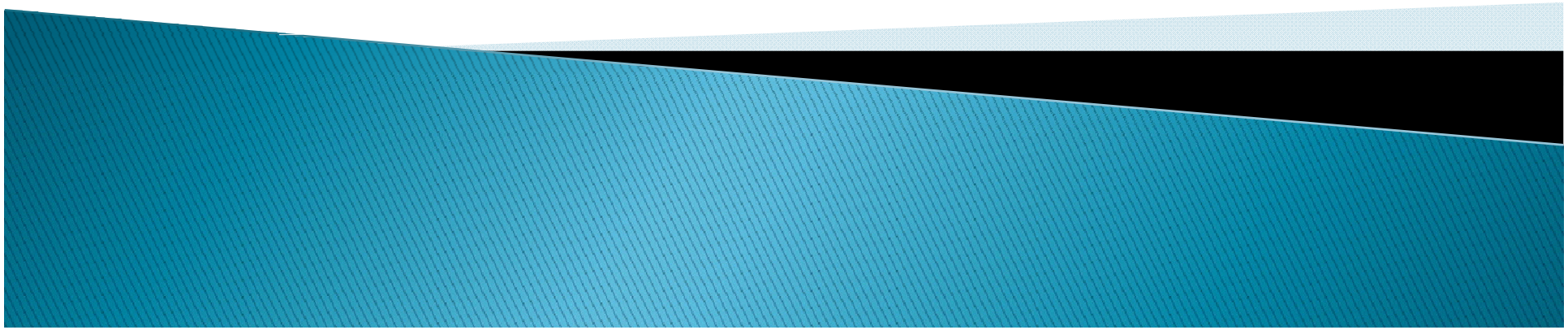
Some Thoughts on Cloud Storage Security based on Tahoe-LAFS

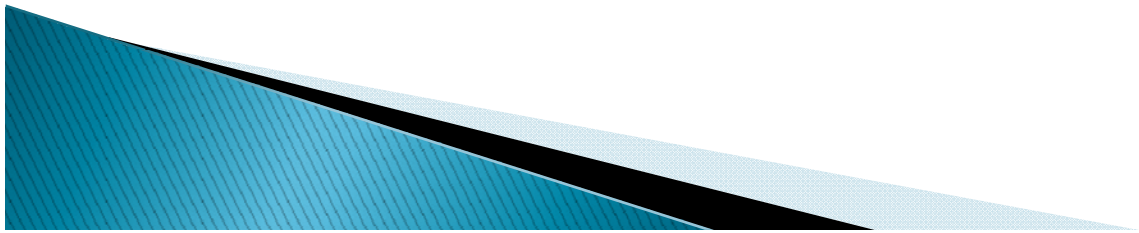
Li Tang

NSLab Seminar, Nov. 2009



Outline

- Cloud Storage – A New Paradigm
- Security Advantage and Challenges
- Tahoe-LAFS
- Beyond Confidentiality and Integrity
- Mechanisms against Snoopers



Data Storage Growth

Traditional Data

Documents
Character & numerical databases

+

Additional, New Data

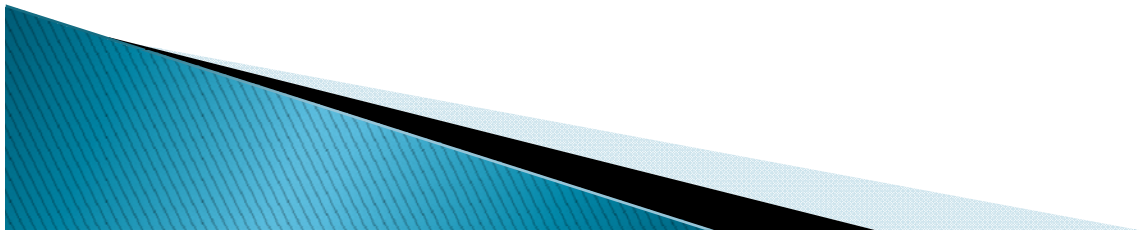
Images – 500KB per picture
Audio – 5,000 KB per song
Video – 5,000,000 KB per movie

Digital Content
- 85% of all data by 2012
- Growing 10x every 4 years

Source: IDC

Cloud Storage Definition

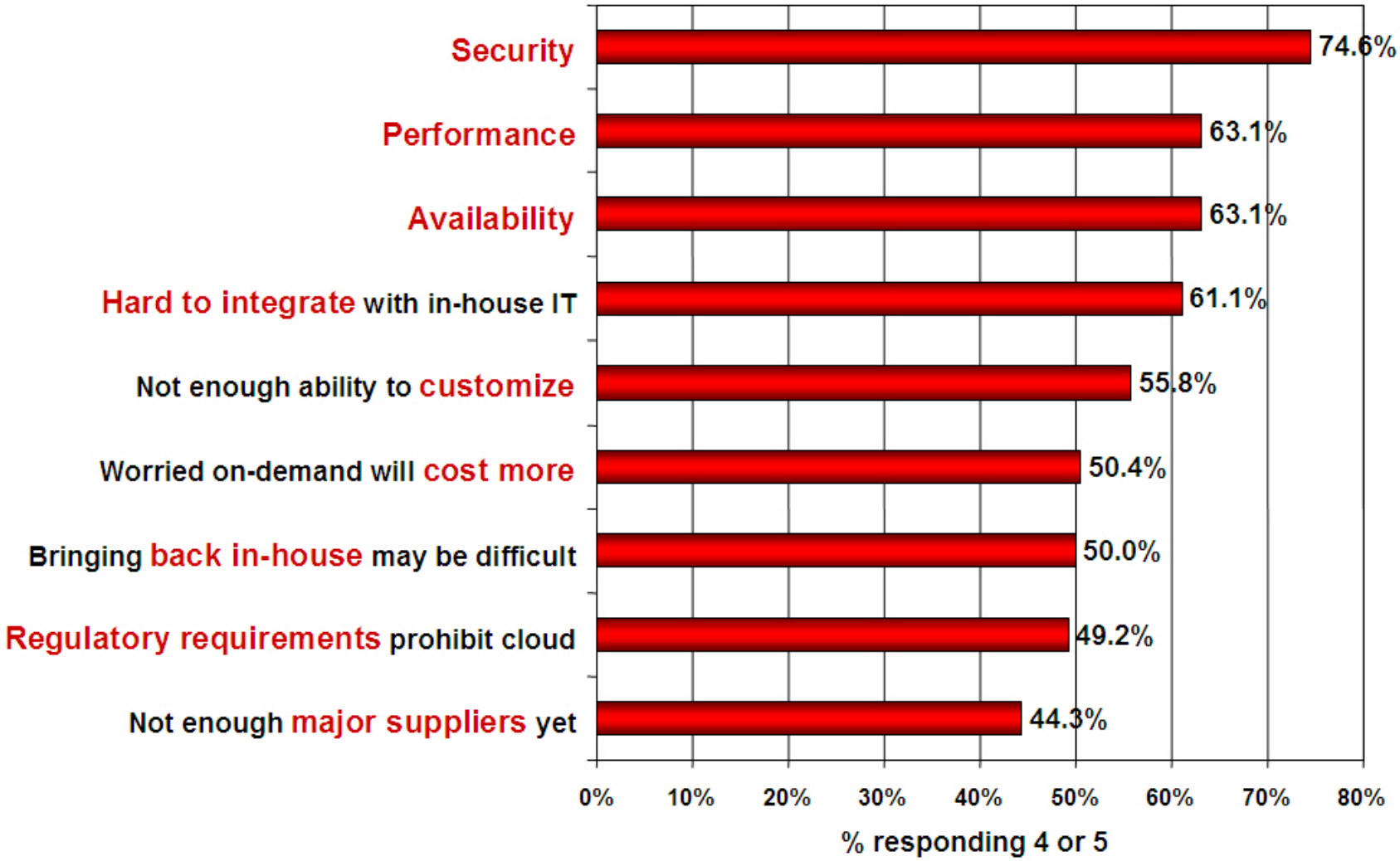
- Cloud Storage Places Data Outside the Walls
  - “Cloud Storage” is defined as storage which resides in a public or private infrastructure that is external to the primary storage infrastructure, and is often shared to some extent
  - Cloud Storage is different from Cloud computing, where a whole application lives fully or partially in the Cloud



Security is the Major Issue

Q: Rate the challenges/issues ascribed to the 'cloud'/on-demand model

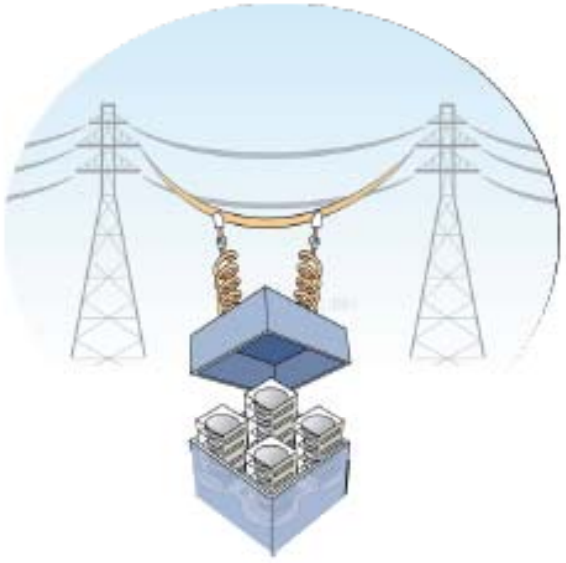
(1=not significant, 5=very significant)



Source: IDC Enterprise Panel, August 2008 n=244

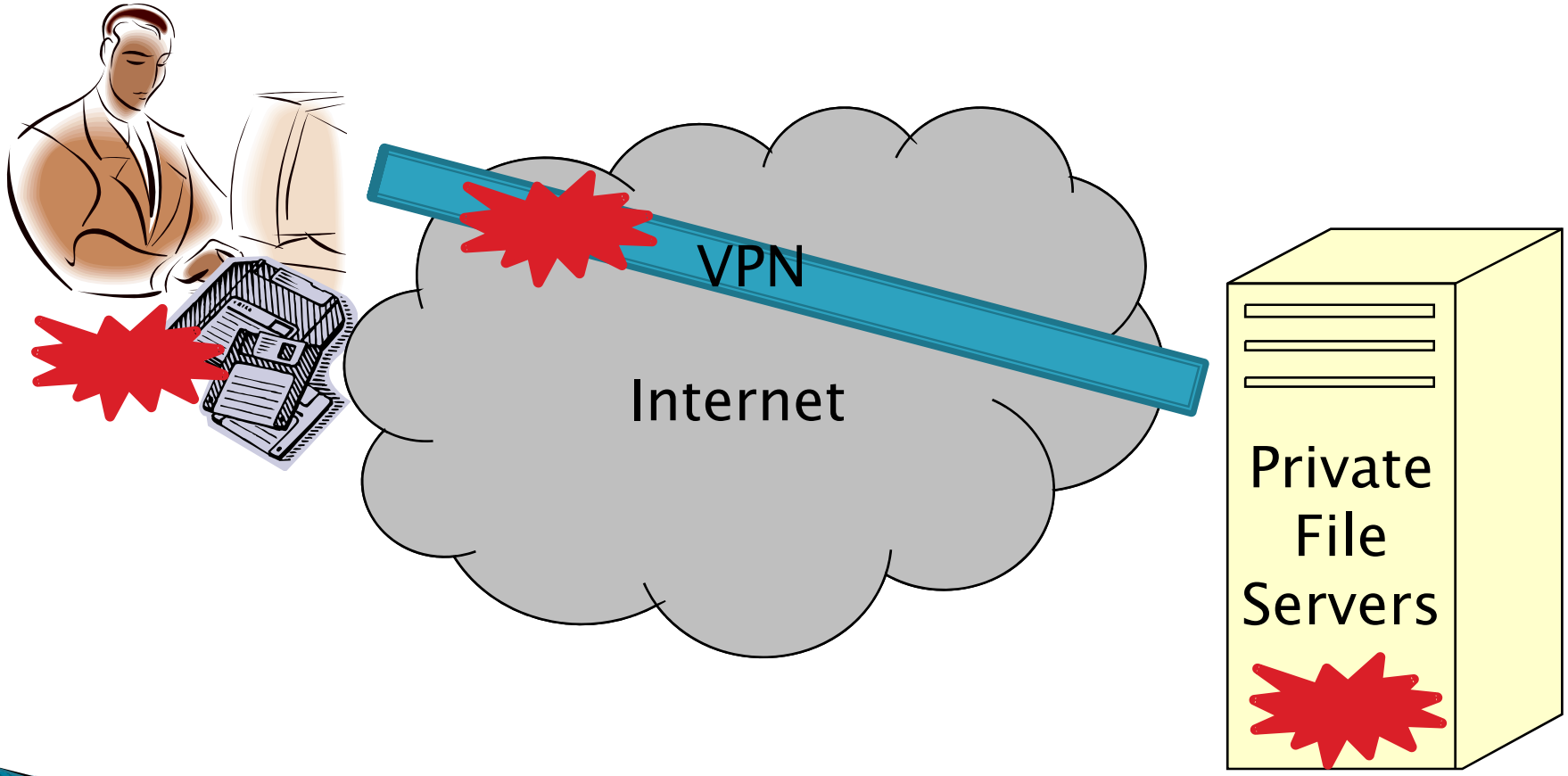
What do Storage Buyers Really Want?

- Storage is more like a bank than a utility
  - How do I know this is secure?
  - How do I get my data back if you belly-up?

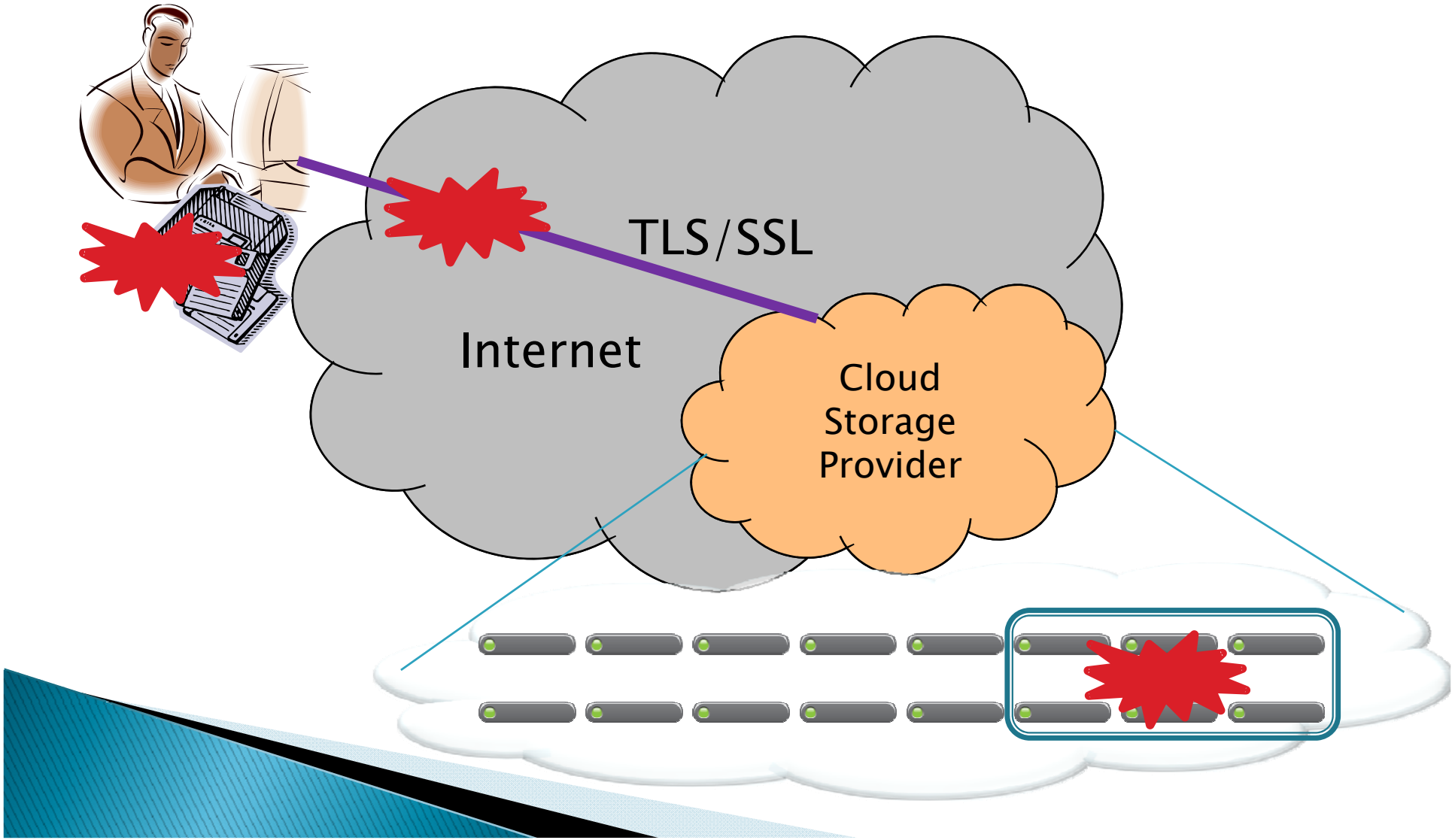


OR

Security Model of Traditional Storage

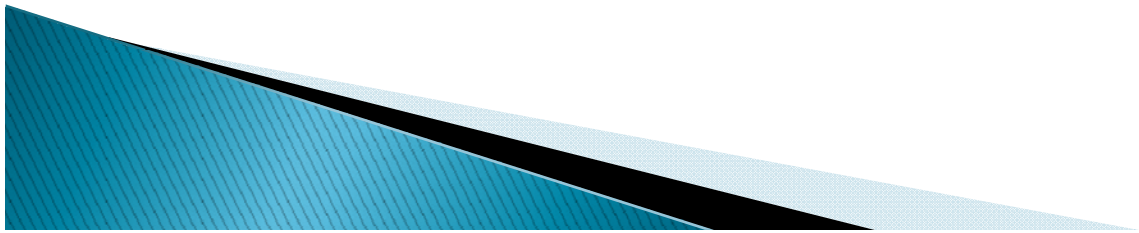


Security Model of Cloud Storage



High-level Perspective

- Old Games
  - Protect data presented to owners
  - Protect data stored on servers
  - Protect data communication
- Fundamentally New
  - Relationship between data's owner and holder
  - Traditional data storage
    - The same, or complete trust
  - Cloud storage
    - Relying on service contract



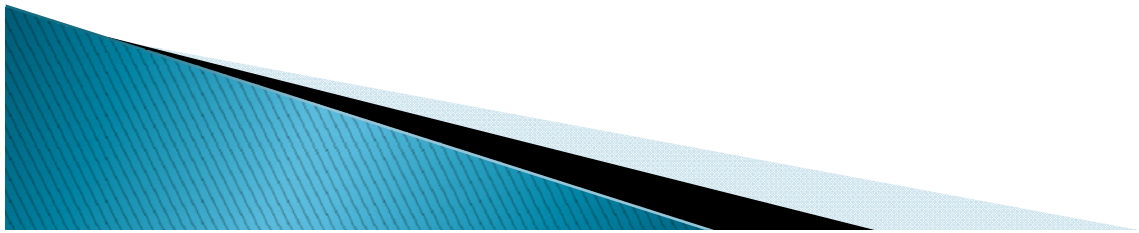
General Advantages and Challenges

- Professional security management
- Homogeneity eases security auditing/testing
- Wide-area backup increases reliability and disaster recovery
- High-class infrastructure enhances availability

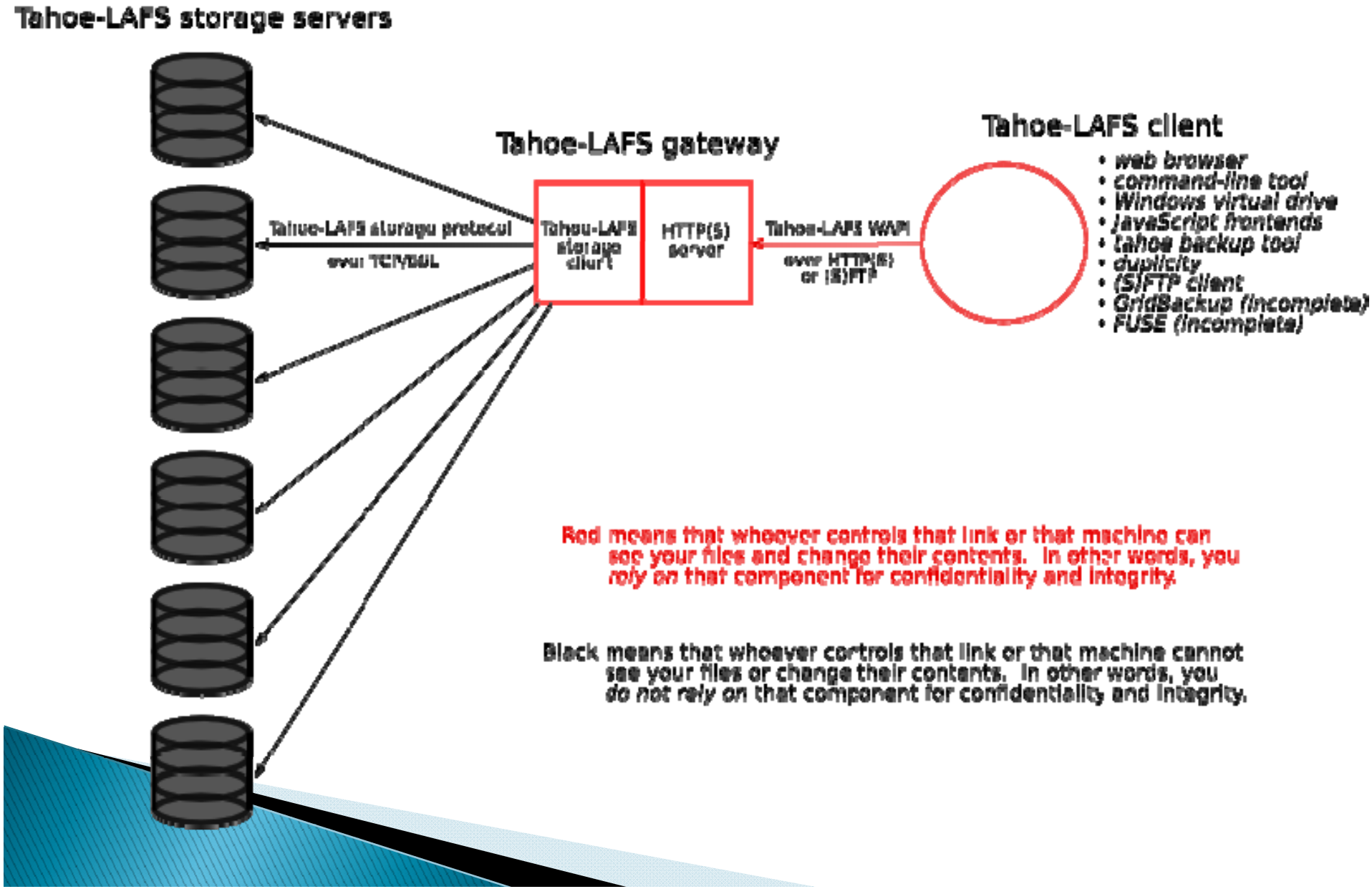
- Loss of physical control
- High dependence on cloud storage providers

Tahoe – A Least Authority File System

- Architecture – provider-independent security
  - Data originates at the client, which is trusted
  - Client segments, encrypts, and erasure-codes data
  - Segments are distributed to storage nodes over secure links
  - Storage nodes, which are not trusted, only see encrypted data
- Latest Status
  - Open Source, Release 1.5.0
  - Sponsored by AllMyData.com
  - Included in Ubuntu Karmic Koala

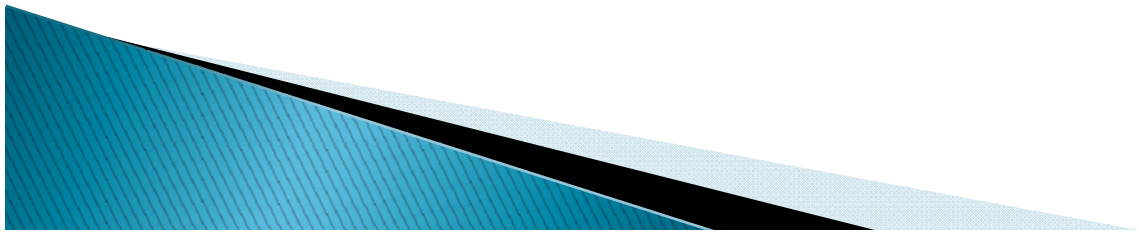


Tahoe-LAFS network topology

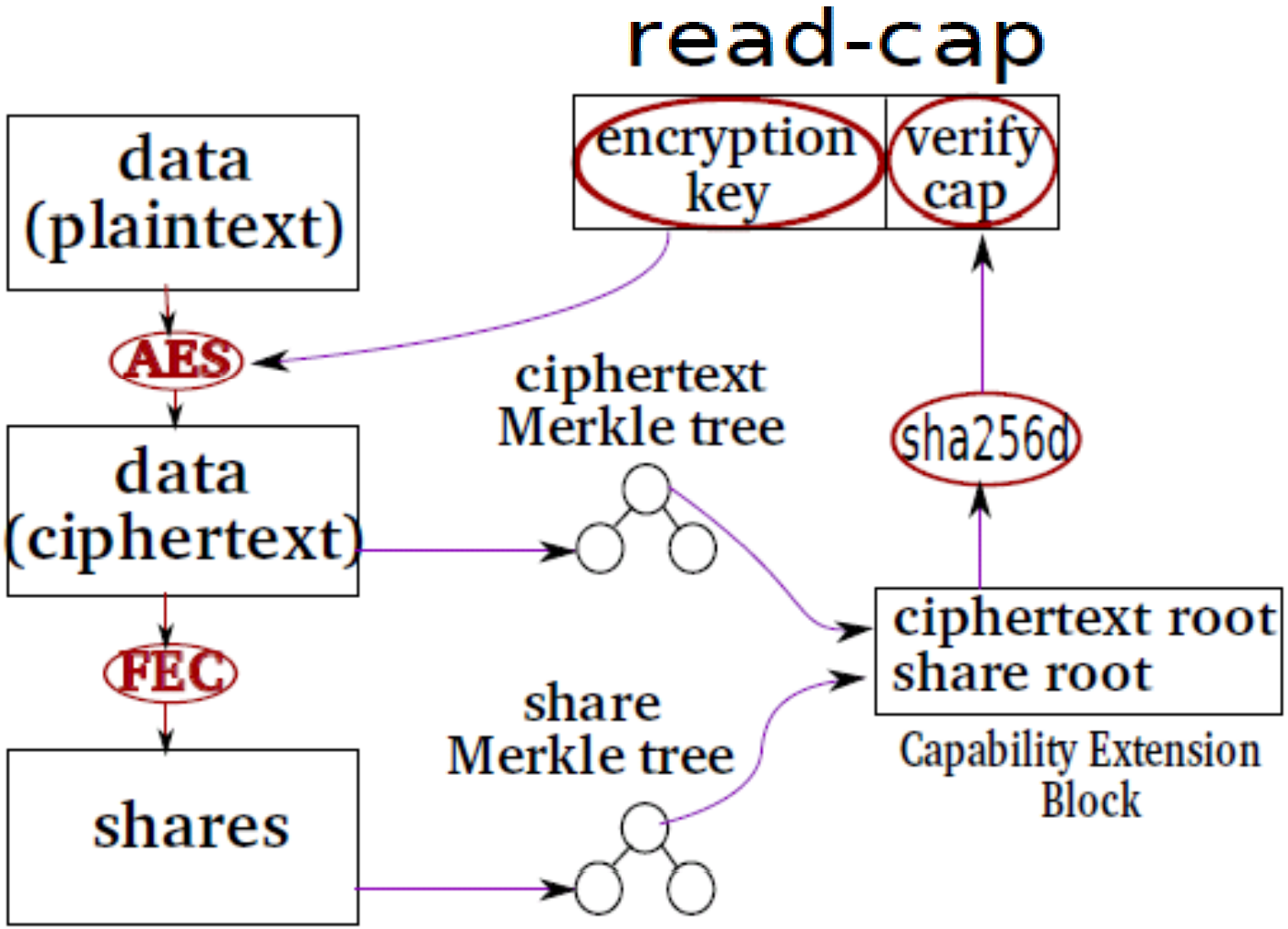


Access Control

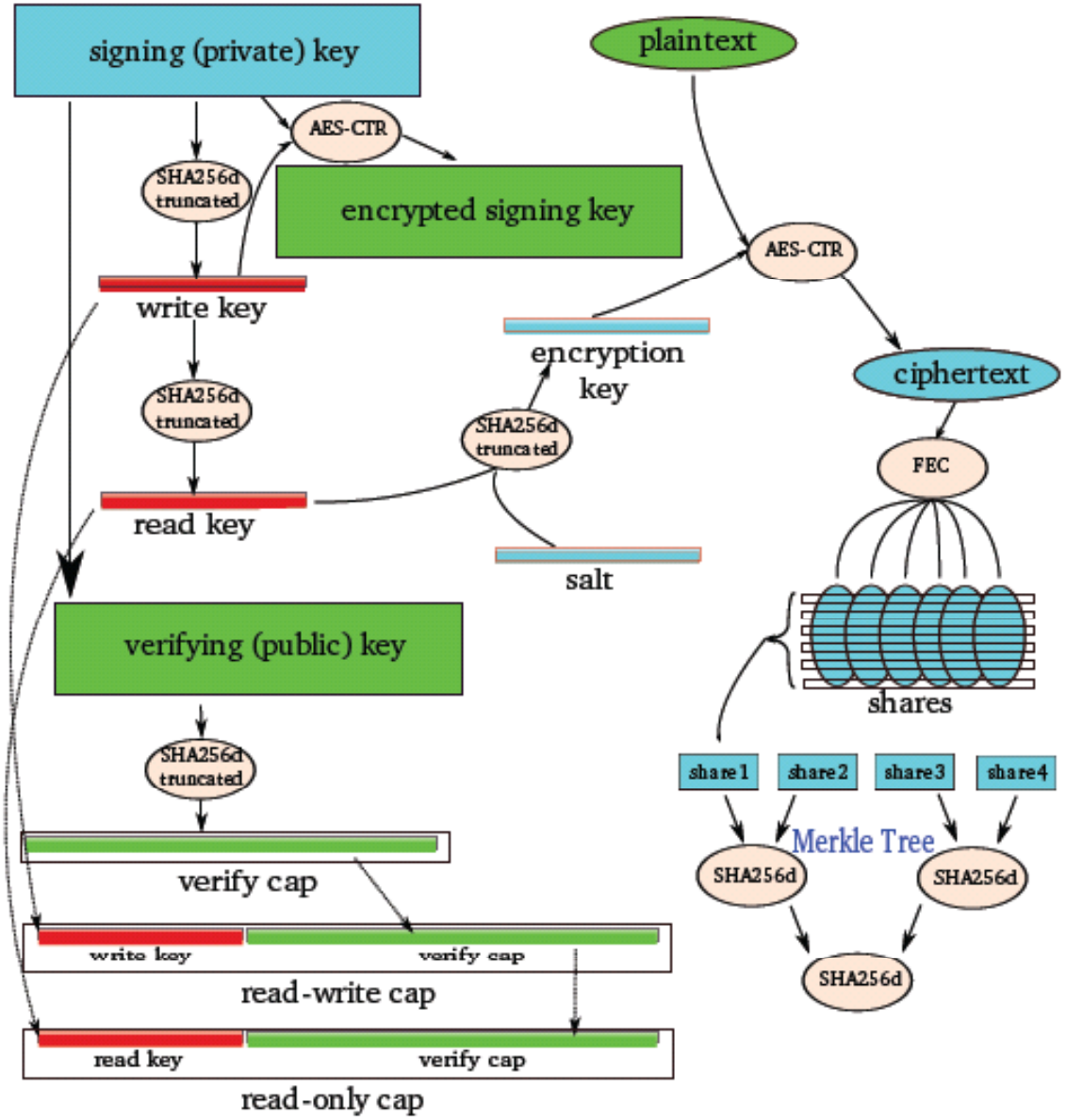
- Two types of files
  - Immutable files
  - Mutable files
- Three Classes of Privileges (capabilities)
  - Read-Write-Cap (only for mutable files)
  - Read-Cap
  - Verify-Cap
  - Capabilities are inclusive and self-authenticating
  - Example:
    - URI:CHK:6hwdguhr5dvgte3qhosev7zszq:lgi66a5s6gchcu4yy aji3blogdxmrrrgcdxj5q33bz7h2dhlp6oq:3:10:8448



Immutable File



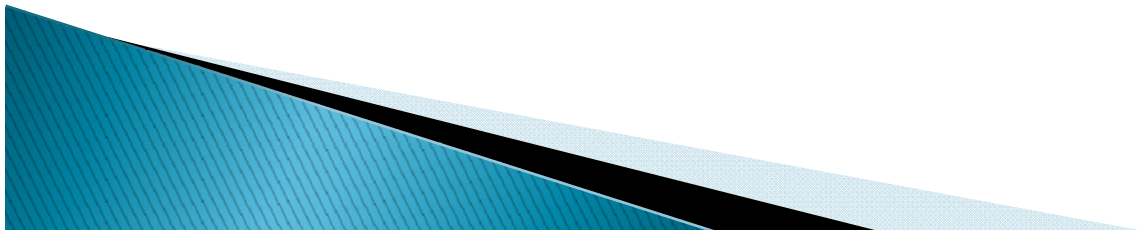
Mutable File



Content Hash Key

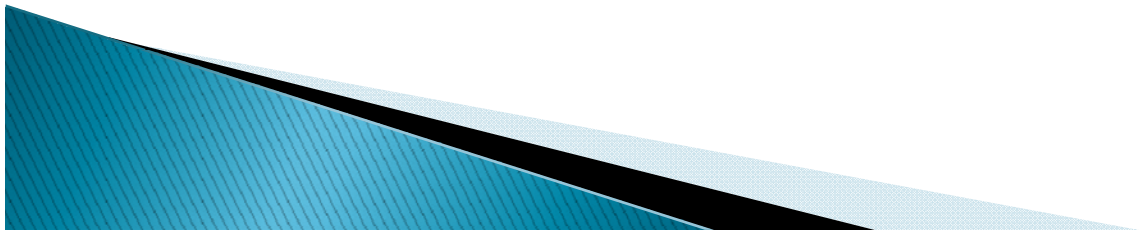
- Key = Hash(Content)
- Advantage – Convergence
  - Plaintext A = Plaintext B

  - Cyphertext A = Cyphertext B
- De-duplication



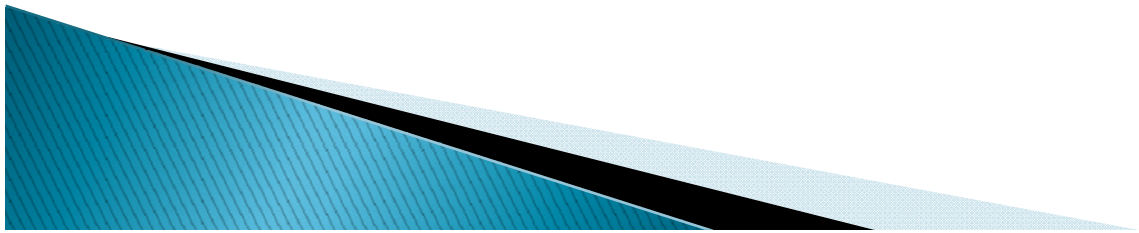
Security Model of Tahoe

- Ensure
  - Confidentiality
  - Integrity
- Not offer
  - Privacy
  - Anonymity
- A Snooping Example:
  - Whether a colleague has filled a medical record that has a standard template



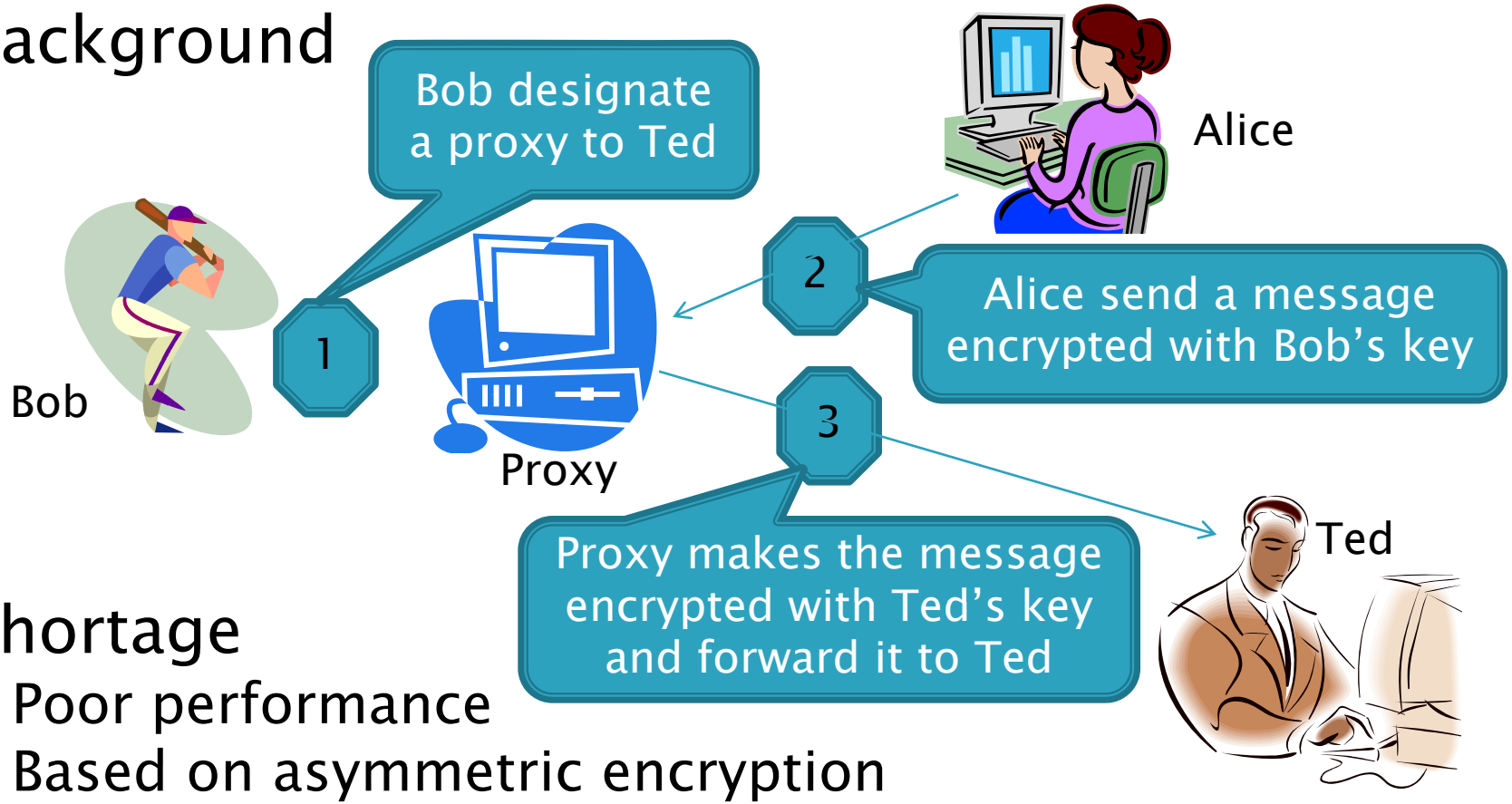
Mechanisms against Snoopers

- Assumption
  - Data holders neither play the role of nor collude with snoopers
- Scheme A
  - Proxy re-encryption
- Scheme B
  - URI randomization



Proxy Re-encryption

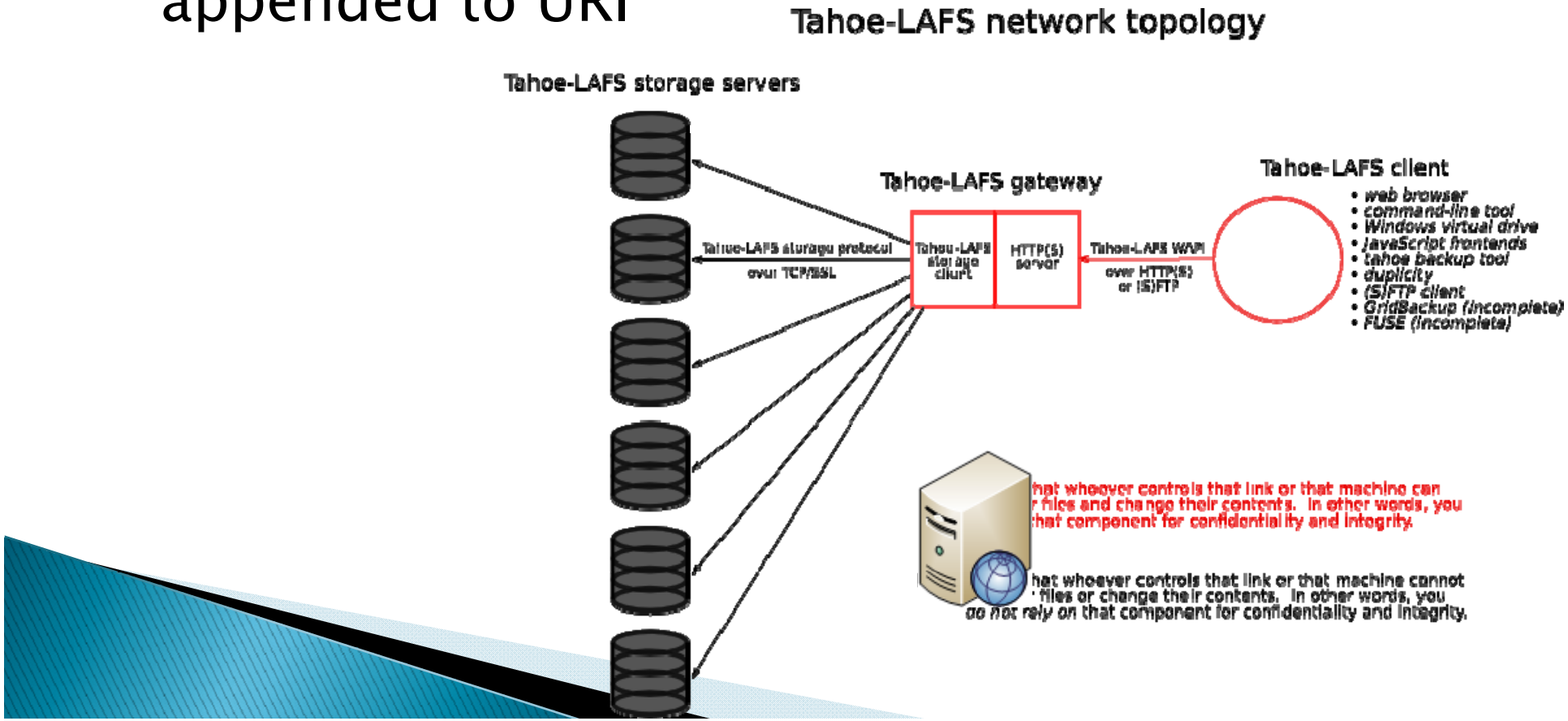
- Background



- Shortage
  - Poor performance
  - Based on asymmetric encryption
  - Complex key management

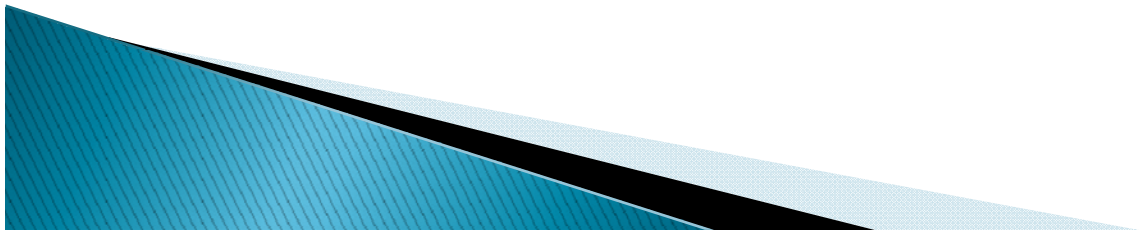
URI Randomization

- Each storage server signs a successful write
- K write signatures barter from the management server a read ticket, which is appended to URI



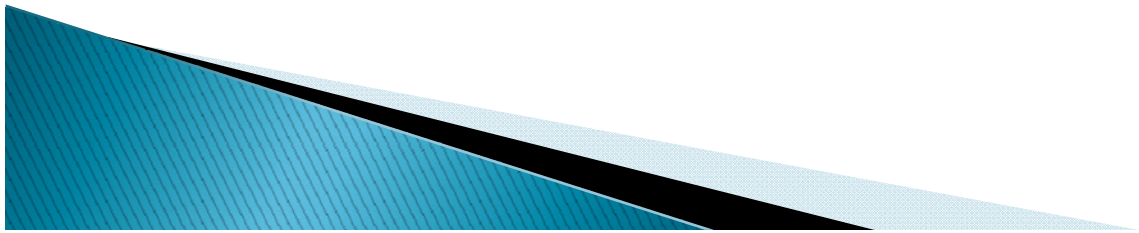
Questions?

Thanks!



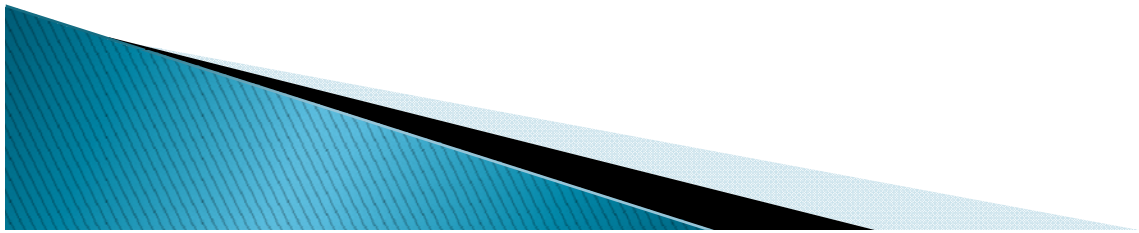
Data Types for Cloud Storage

- Larger files with lots of read access
  - Digital content
- Parallel streaming writes
  - video surveillance (private clouds)
- Long-term storage files
  - Backup and archival files (private clouds)
  - Medical images, Energy exploration, Genomics
- Geographically shared files
  - Access from different geographies (public clouds)
  - Movie trailers, training videos

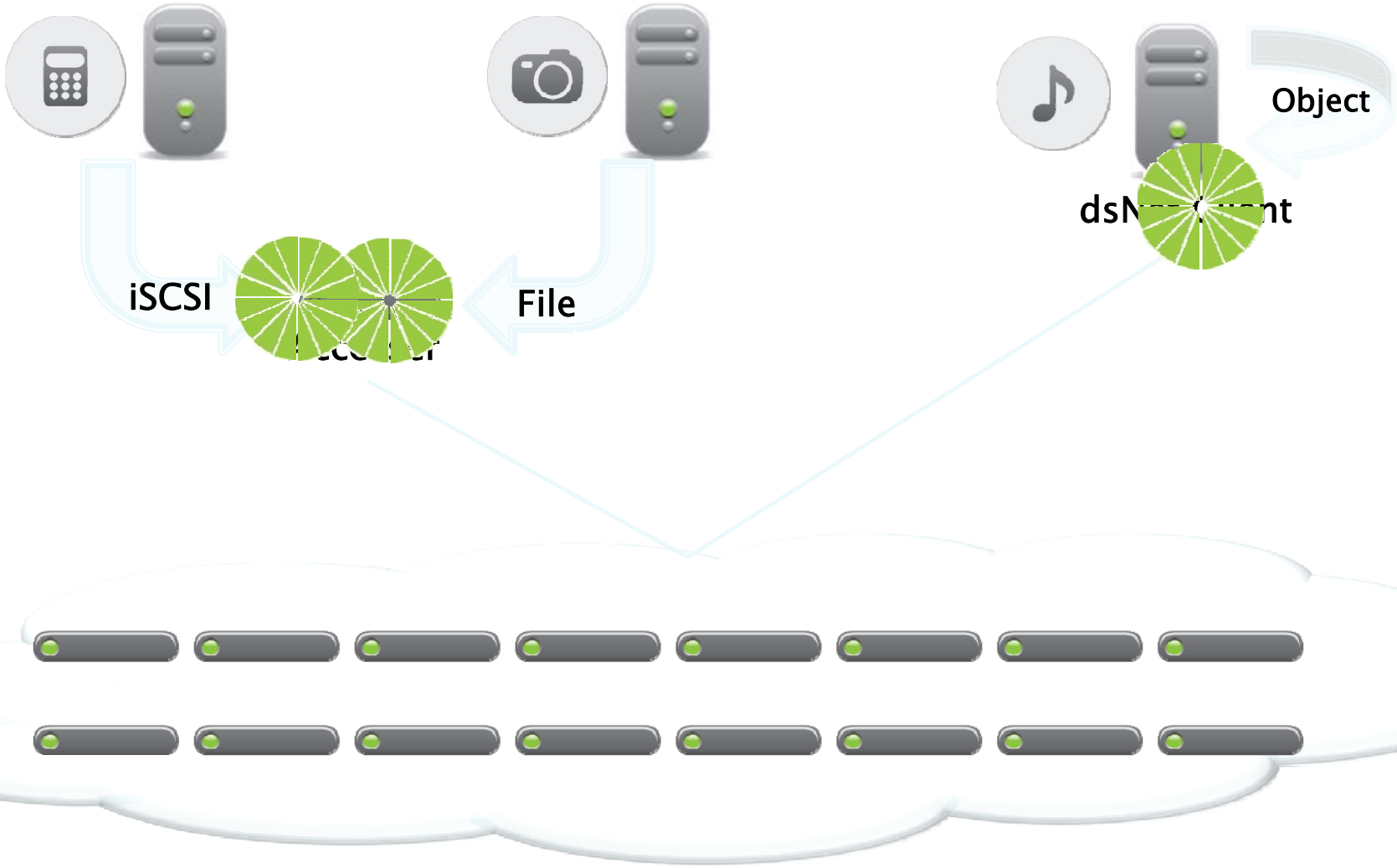


Where is Cloud Storage a Poor Fit?

- Active Corporate Data
  - Advanced data protection schemes
  - Office Documents, Spreadsheets
  - Source-code
- Transactional Data
  - Frequent read and write access
  - Massive I/O requirements
  - Database, source code, Active VMware images



Information Dispersal



Storage Overhead	15-60%	Maximum Delivery	16 at once
Bandwidth Needed	15-60%	Delivery choices	thousands

Cleversafe