Secure Storage Systems and Key Technologies

Jiwu Shu, Zhirong Shen, Wei Xue, Yingxun Fu

Tsinghua University, China

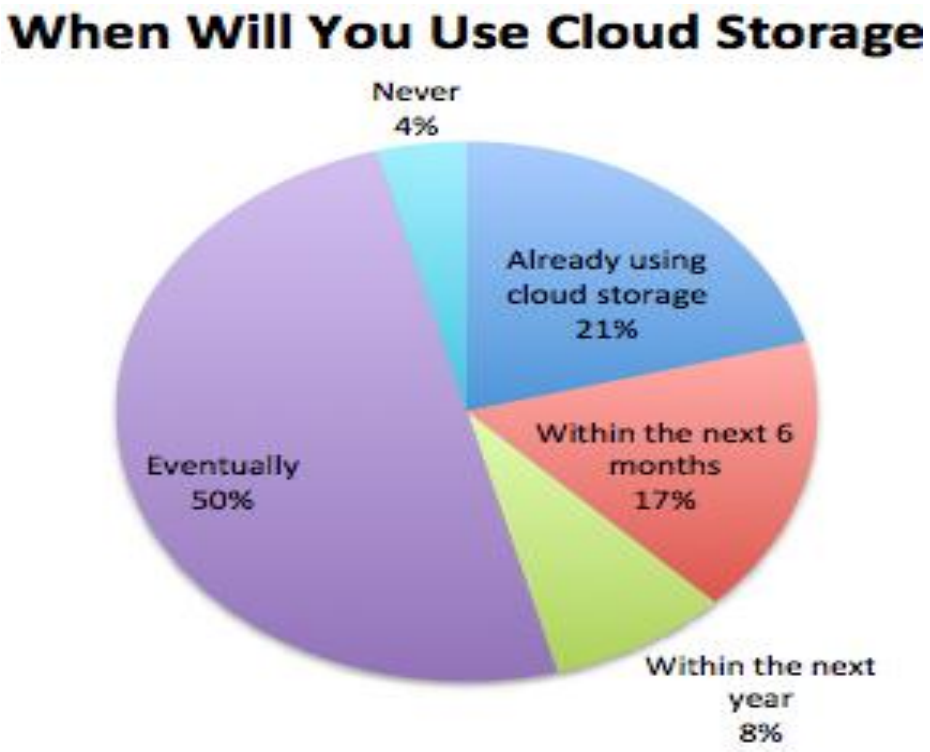
Outline

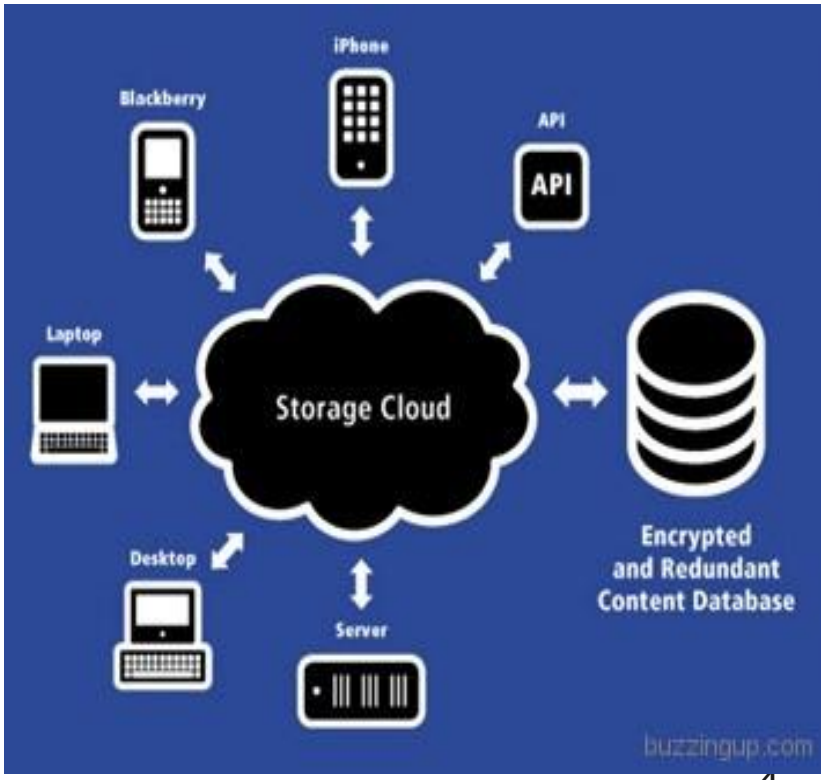
- Research Background
- Design Criteria of Secure Storage
- Basic Technologies
- Extended Technologies

Outline

- **Research Background**
- Design Criteria of Secure Storage
- Basic Technologies
- Extended Technologies

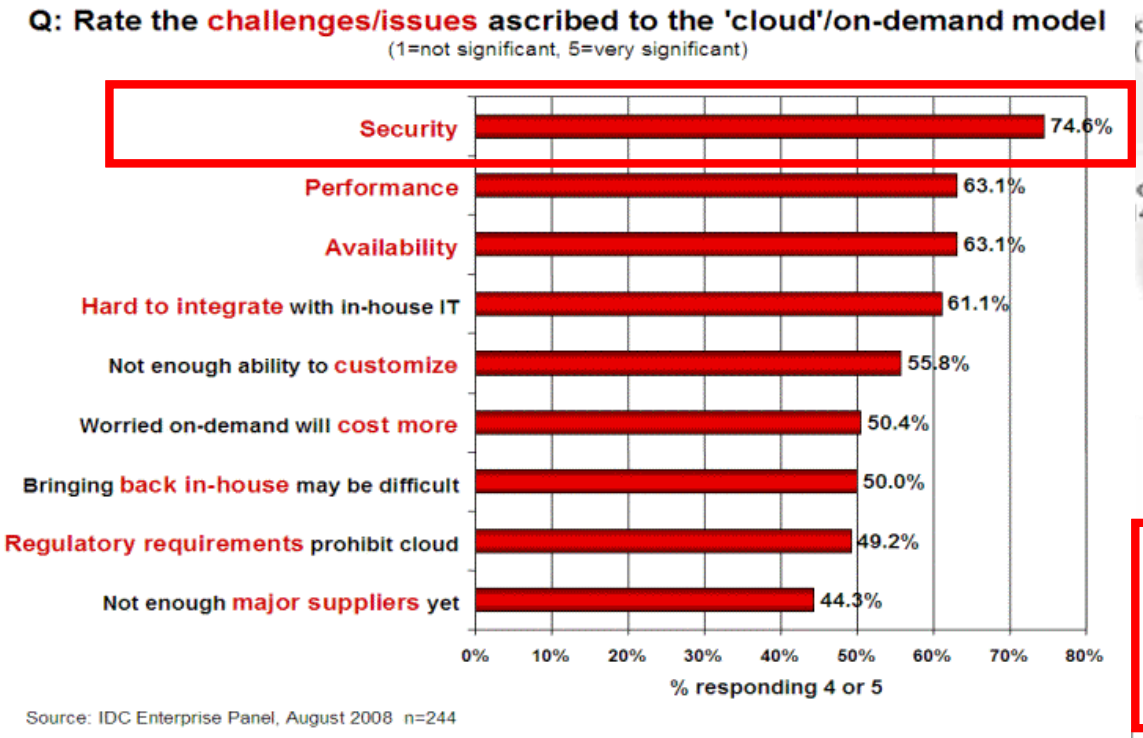
Research Background

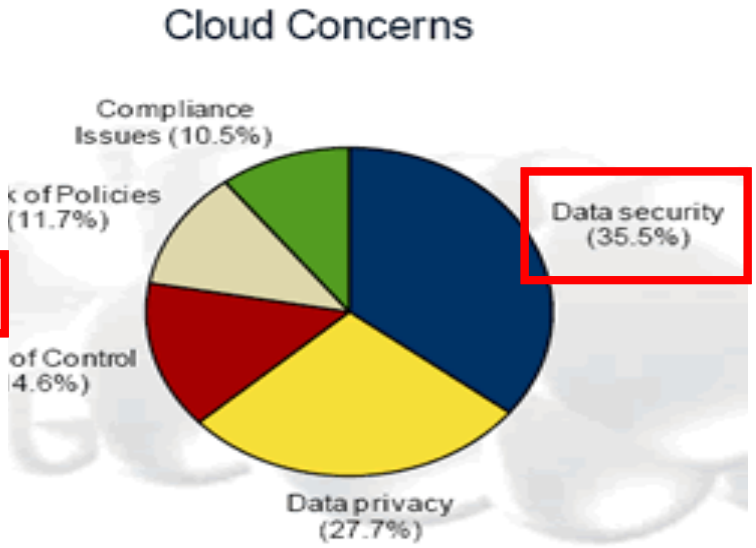




Cloud storage service is widely adopted.

Research Background





Concerns of data security in the cloud are aroused !!

Research Background

How to securely store data in the cloud?

Outline

- Research Background
- **Design Criteria of Secure Storage**
- Basic Technologies
- Extended Technologies

Design Criteria of Secure Storage

- **Confidentiality**
  - Data information are secret against the unauthorized access
- **Integrity**
  - Unpermitted Modification Prevention
  - Unpermitted Modification Detection
- **Availability**
  - An authorized user can execute a data operation within an acceptable period of time
- **Performance and Others**

Outline

- Research Background
- Design Criteria of Secure Storage
- **Basic Technologies**
- Extended Technologies

Basic Technologies- Secret Key Distribution

Data are encrypted, how to distribute keys?

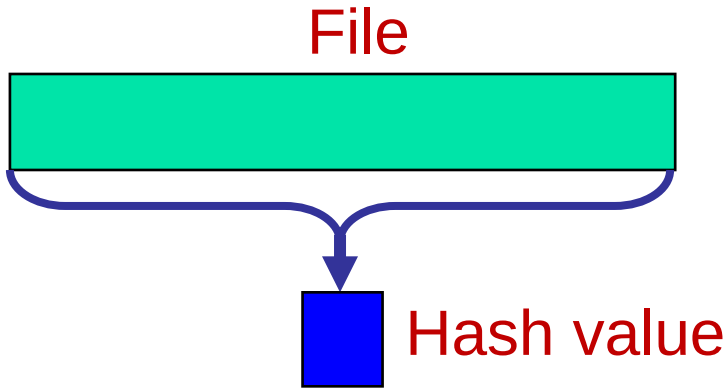
- **Servers-dominated schemes**
  - Need to fully trust servers
  - Inadequate when servers are untrusted
- **Owner-dominated schemes**
  - Owners should always online
  - Incur huge management burden
- **Trusted-third-party-dominated schemes**
  - Avoid putting complete trust on servers
  - Save data owners from burdensome management

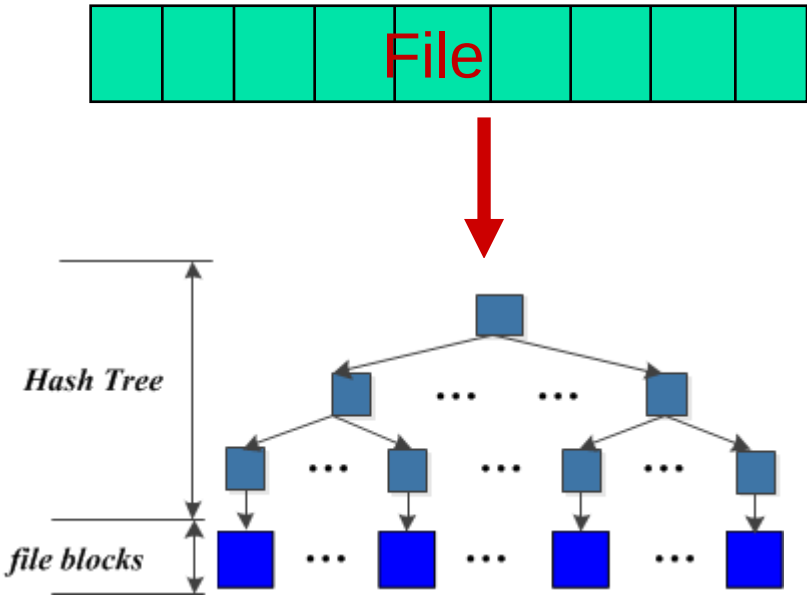
Basic Technologies- Access Control

- **Public/private keys**
  - Writer- private key, reader- public key
  - Asymmetric operations, cost expensive
- **ACL + symmetric keys**
  - Writer gets more keys to append signature
  - ACL recodes all users and is stored in metadata file
- **ACL + public/private keys**
  - Employ user's public key to encrypt file block key
  - Reduce the amount of keys to manage for clients
  - E.g., ACL only recodes the writers in FARSITE

Basic Technologies- Integrity Checking

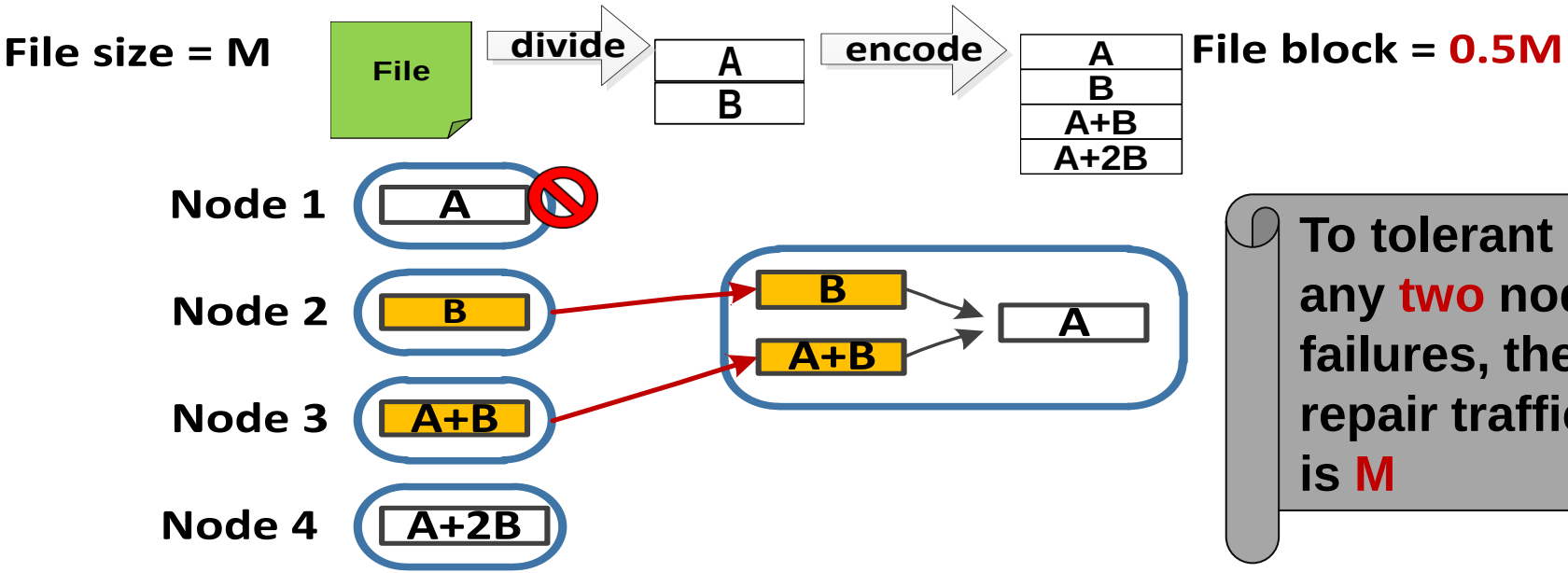
- **Calculate the hash value H_i for file F_i**
  - Convenient but not efficient
  - Every access demands the re-calculation of H
- **Construct Hash Tree**
  - Partition the files into many file blocks
  - Construct Hash Tree based on the integrity information of file blocks
  - Access/Update Complexity is O(log N)





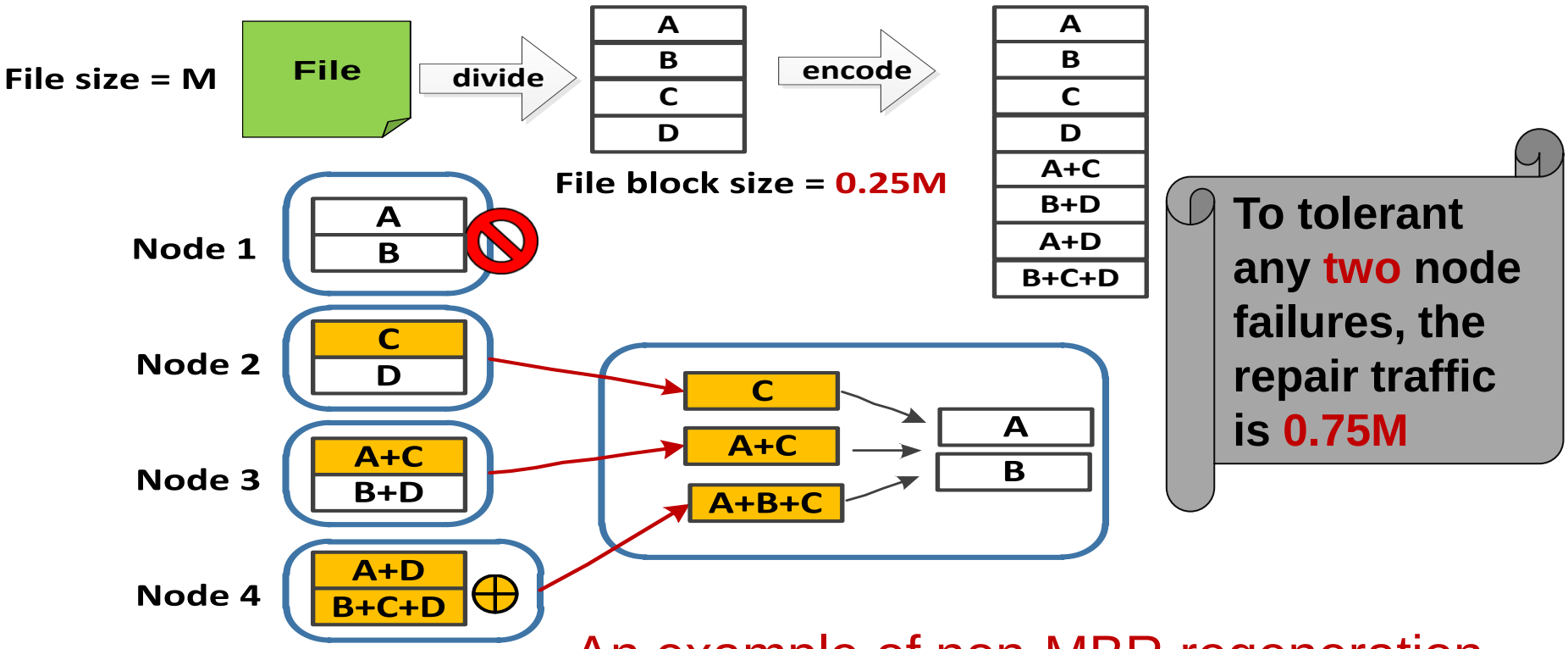
Basic Technologies- Availability Assurance

- **Replication**
  - Storage Cost: tM, Repair Traffic: *M*
  - Fault tolerance: any (t-1) failures of files
- **(n-k) Erasure-Code**
  - Storage Cost: nM/k, Repair Traffic: *M*
  - Fault tolerance: any k failures of file blocks



Basic Technologies- Availability Assurance

- **Regenerating Codes (MBR)**
  - Storage Cost = Repair Traffic = [(2n-2)M]/[(2n-k-1)k]
  - Fault tolerance: any k failures of file blocks
  - Balance storage cost and network traffic

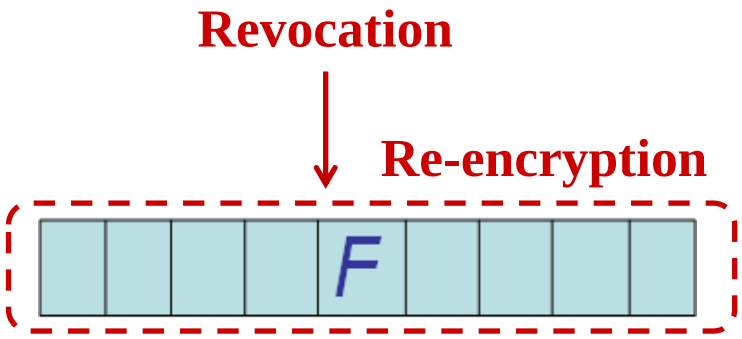


An example of non-MBR regeneration

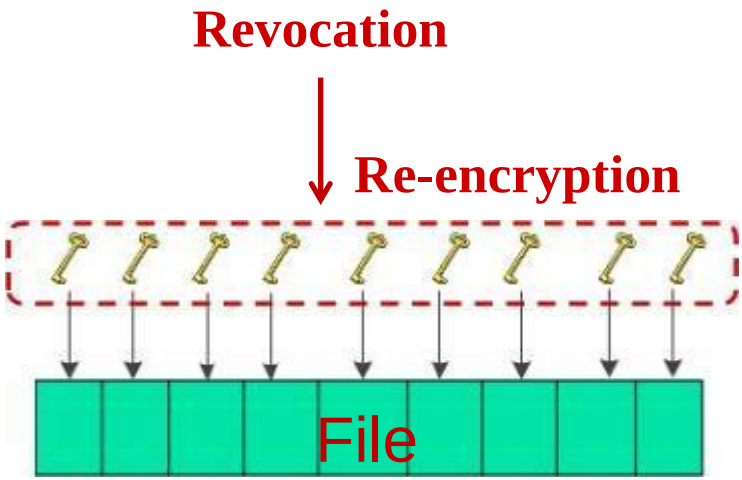
Basic Technologies- Permission Revocation

When a user's permission is revoked…

- **Aggressive revocation**
  - Regenerate new keys and perform re-encryption for the involved files
  - Timely but expensive



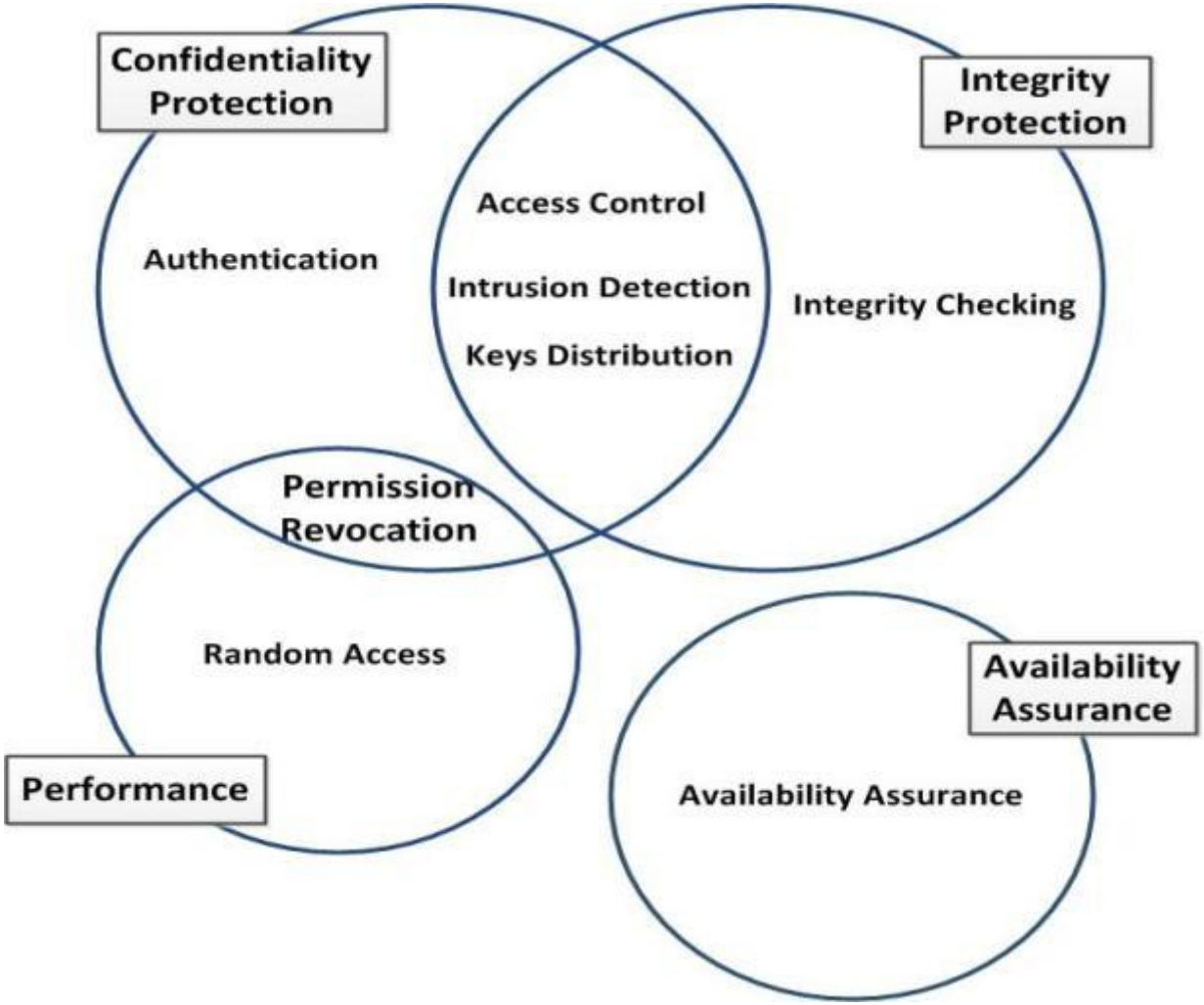
- **Lazy Revocation**
  - The revoked user can still read the unchanged files after the revocation
  - Defer the re-encryption to the update of the involved files
  - Complex management: e.g., key versions



Basic Technologies- Others

- **Authentication**
  - Authentication is the first guard to prevent unauthorized user's illegal operations
  - Authentication in cloud storage system can simply classify as public-key-infrastructure-based and user ID with authentication-key-pairs-based authentication
- **Storage-based Intrusion Detection System**
  - Storage-based intrusion detection system (SIDS) is a significant part of intrusion detection system (IDS)
  - SIDS usually deploy in storage device
  - SIDS is usually workable even if the host is invaded, it is very suitable for cloud storage server

Basic Technologies- Summary



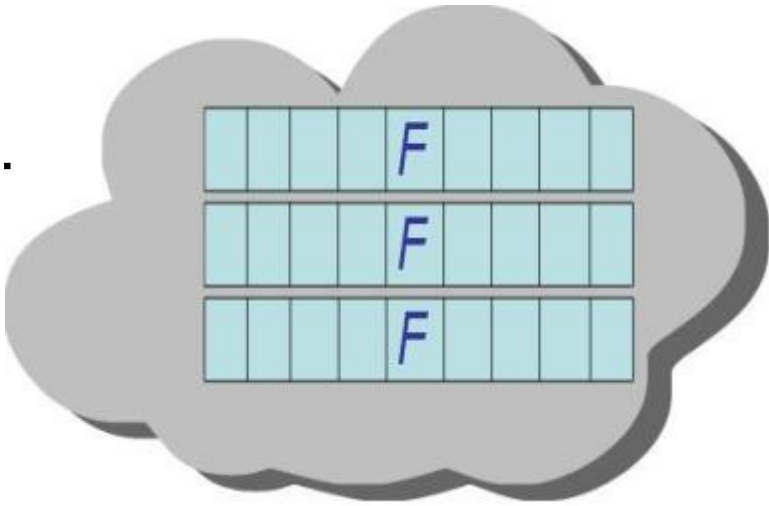
Outline

- Research Background
- Design Criteria of Secure Storage
- Basic Technologies
- **Extended Technologies**

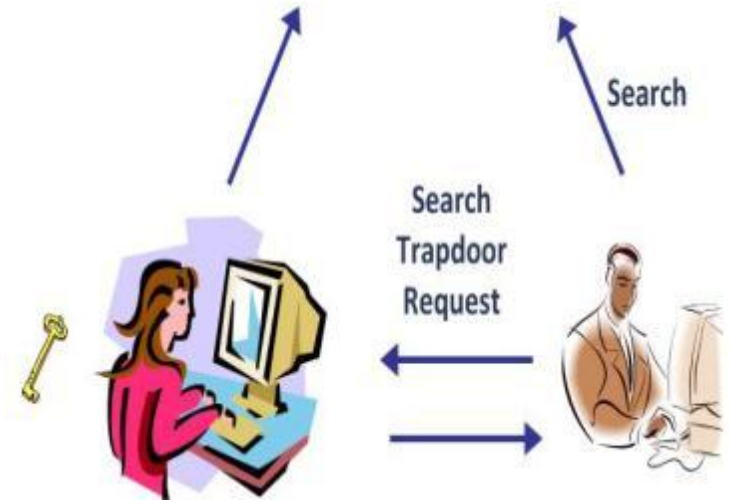
Extended Technologies- Searchable Encryption

Data are stored in encrypted form, when a user plans to find some files…

- **A Straight Forward Way**
  - Download and decrypt all the files
  - Perform search on plaintext
  - Huge cost for unneeded files



- **Searchable Encryption**
  - Users ask owner for search trapdoor
  - Cloud server performs search based on the trapdoor
  - Efficient and secure



File Sharing

Extended Technologies- Searchable Encryption

- **Research status of Searchable Encryption**
  - Single keyword search support

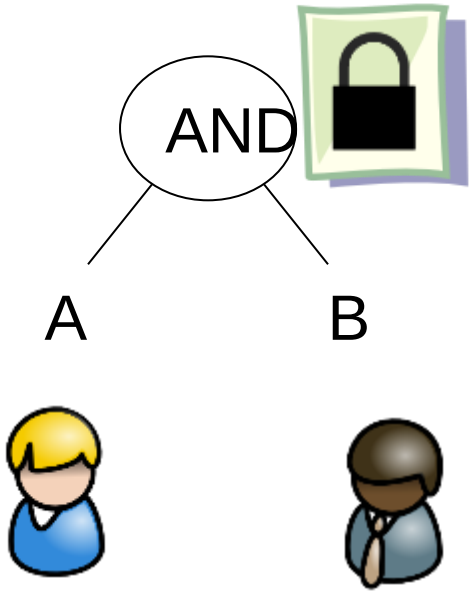
    “conference=A”
  - Conjunctive query support:
    “(name=Alice)and(age=20)”
  - Logical query support over multiple keyword fields:
    “(name=Alice or Bob)and(age=20or16)”
  - Ranked keyword search

    return the top-k relevant results
  - Similar keyword search

    “*lice”→”Alice”,”Blice”,…,”Zlice”,”lice”

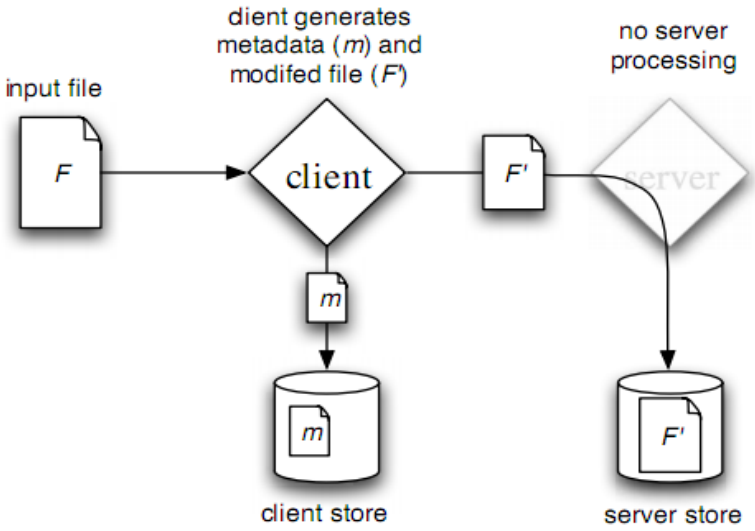
Extended Technologies- Attribute-based Encryption

- **Motivation**
  - Every user has to manage massive keys in the cloud
  - Neither flexible (number of keys), nor convenient (online distribute keys)
- **Use attributes as the keys**
  - Attribute: age, profession, etc
  - Flexible (only attribute keys), convenient (offline access)
  - Flexible access control mechanism: ‘and’, ‘or’, ‘in’ gates

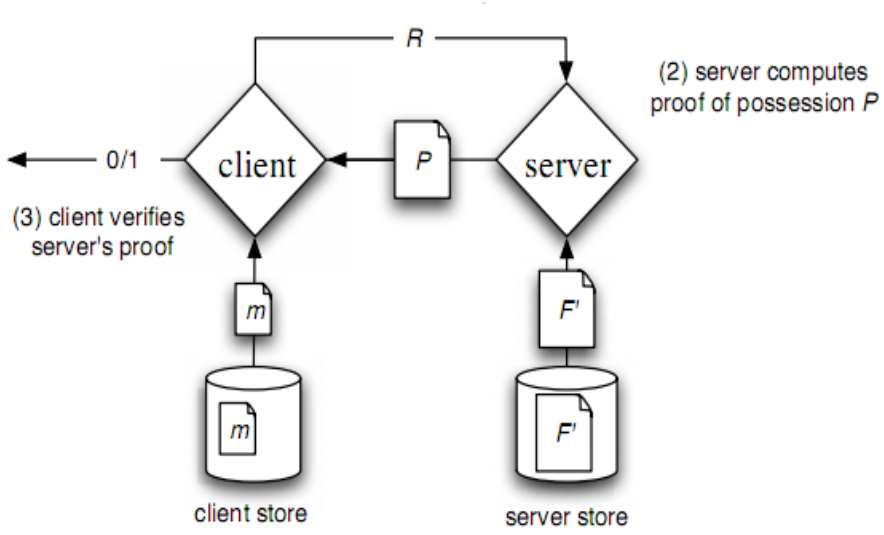


Extended Technologies- Data Audit

- **Provable Data Possession (PDP)**
  - PDP scheme enable users to checking outsourced data's integrity with low computation and communication cost
  - PDP can category as basic PDP, DPDP, and MR-PDP
  - Use Third-Part Auditor (TPA) to provide public verifiability and batch verifiability mechanism



(a) Pre-process and store



(b) Verify server possession

Extended Technologies- Data Audit

- **Proof of Retrievability (POR)**
  - POR scheme can simple consider as two parts: verification and recover
  - POR's verification part is assemble as PDP's
  - POR's recover part usually utilize reliability technologies, such as replication, erase code, network code, and so on
- **Data Provenance Auditing**
  - Provenance can be consider as the information that helps cloud storage to describe the derivation history of data
  - Auditing provenance could analysis the origin of the cloud data, which is crucial in multiple user cooperative work system

Extended Technologies- Data Assured Deletion

- **Data Assured Deletion**
  - Assured Deletion is that cloud data would become completely unretrievability after the certain condition
  - Fade utilizing a third party to manage keys and provide policy-based assured deletion mechanism
  - Vanish utilizing the special feature of DHT network, implement a time-based assured delete mechanism on the email scenario
- **Secure Data Deduplication**
  - Traditional data deduplication mechanisms may cause some security problems
  - Data encryption has increased the difficulty of data deduplication

Thank you!