A peer-to-peer based file backup system

Student
Alessandro Duminuco
Supervisor
Prof. Ernst Biersack

EURECOM
Sophia Antipolis

University
Institut Eurécom
Sophia Antipolis
France

General Objectives & Assumptions

Backup Systems

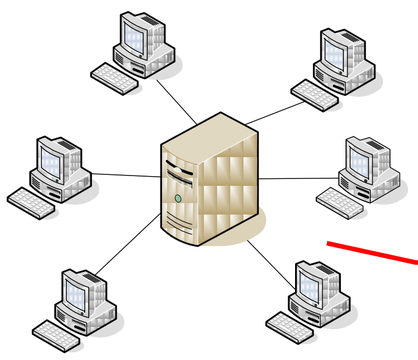
It is easy to produce big amount of digital data.
Most of them are extremely important.
Digital data storage is intrinsically unreliable.

Backup systems store copies of important data with very high reliability.

Central Backup Systems

In companies and universities data backup strategy is applied to the whole network.

Traditional approaches consist of centralized servers collecting important data and taking care of their durability.



Problems
- Single point of failure
- High Cost

Peer-to-Peer Backup Systems

- Every peer in the network can ask the system to backup data.
- Data is spread among peers themselves exploiting their spare disk space.
- If redundancy is added to data, they can be restored in any moment from alive peers.

Peer-to-Peer Systems

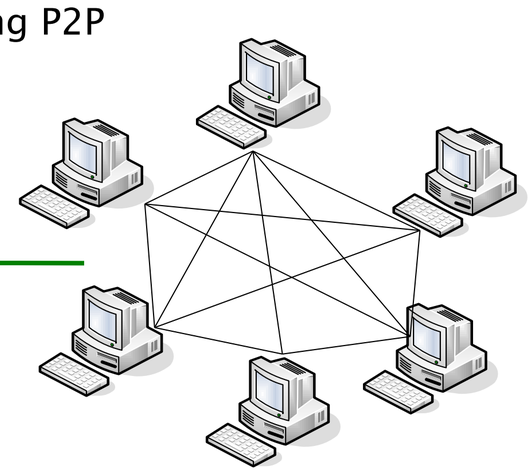
Different way to organize computers in a network.

Services are not provided by a server, but by peers themselves in a decentralized fashion. Every peer contacts directly other peers to obtain the service, while providing it to them.

Resources increase along with demand.

Example of existing P2P systems

- File sharing
- Content distribution
- Voice over IP



Advantages
- Scalability
- Low Cost
- Self organizing

Target Environment

- A company or academic network with few hundreds of machines (100-500).
- Machines are equipped with big hard disks (~100GB).
- Are heterogeneous in performance and reliability.
- Are connected through a LAN (10-100 Mbps).
- Part of machines are connected most of the time while others join and leave the network quite often.

Design Objectives

- Low Administration Effort
- Low Traffic Generation
- Low Storage Consumption
- High Level of Decentralization

Requirements & constraints

- Snapshot service: Every user can backup a whole file tree.
- Reliability and persistence: Backed up files are never lost.
- Availability: File restoration may require a bit of time.
- Capacity: ~10 GB per user.

Main Issues & Challenges

Confidentiality

Files of a user are stored on other peers.
How to protect their confidentiality?

Space Efficiency

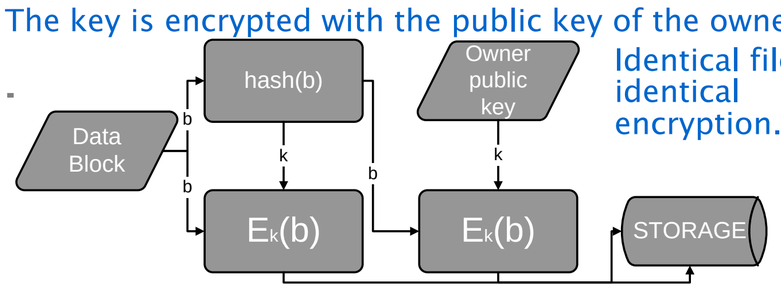
Many users own similar files (e.g. configuration files).
How to store identical files only once?

Redundancy: Placement and Maintenance

Simple redundancy is not enough to provide data durability.
Peers join and leave the network quite often.
How to place and maintain redundant information to provide data durability and availability?

Fairness

Peers backing up their files should provide in turn backup space.
How to prevent free riders?

Private/public key encryption	Data blocks are encrypted using as key the hash of the block itself The key is encrypted with the public key of the owner. Identical files have identical encryption.
Convergent Encryption	
Randomized Replica Maintenance	Data Block Indexing - DHT Infrastructure Redundancy - Network Coding Randomized Maintenance - Peers poll randomly one another - Extract statistics from collected data - Perform randomized procedures based on statistics
?	To Be Addressed

Status

Problem analysis and literature research	General Architecture Design	Storage Layer Architecture: File Organization	Network Coding Study	Design and testing of a randomized replica maintenance algorithm	?	Solving Network Coding Security Issues	?	Design of Application Layer	?	Development and testing	?

Future Challenge:
GOING ON THE WEB??

Progress Bar