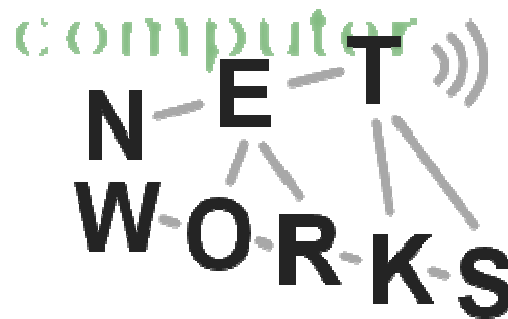


P2P-Networks

Advanced Computer Networks
Summer Semester 2012



Introduction to P2P Systems

- **What?**
 - What is a P2P system?
- **Why?**
 - Why do we need P2P system?
- **How?**
 - How well is the P2P system doing?

What is a P2P System?

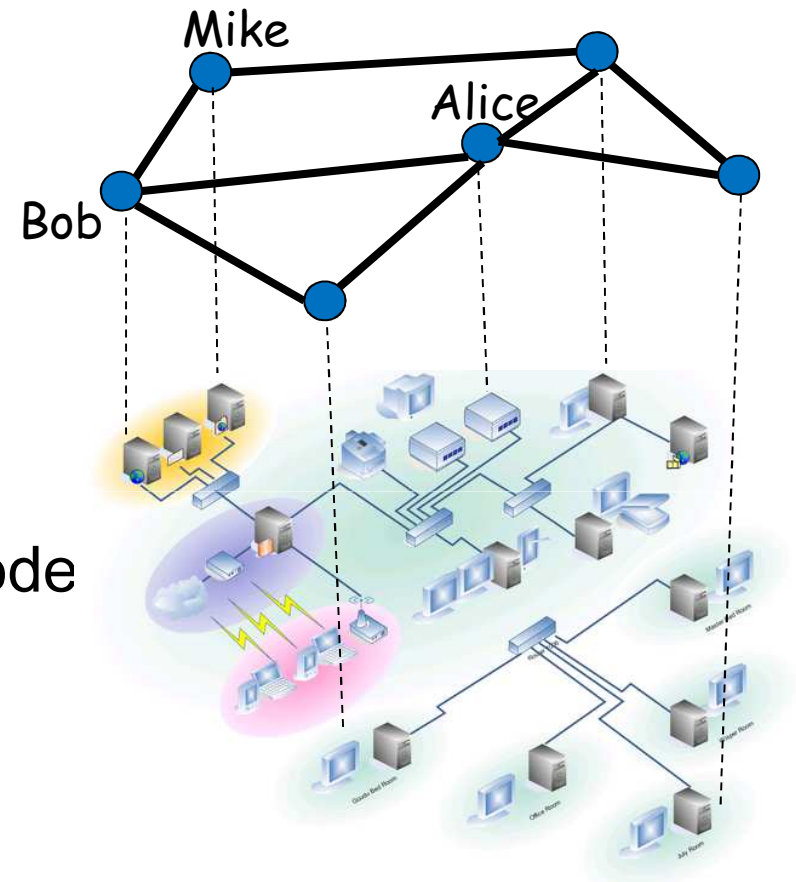
- Peer-to-peer (abbreviated to P2P) refers to a computer network in which each computer in the network can act as a **client or server** for the other computers in the network, allowing shared access to files and peripherals **without the need for a central server** [Wiki]



- The sharing of computer resources by **direct exchange**, rather than requiring the intermediation of a centralized server. [1]
- Their ability to treat instability and variable connectivity as the norm, automatically **adapting to failures** in both network connections and computers, as well as to a transient population of nodes. [1]
 - Fault-tolerant, self-organizing
 - ...

Features

- Decentralized:
 - No central component
- Role: “all peers are equal”
- Self-organized
 - Highly dynamic behavior of node
 - Free to come, free to go
 - Unreliable service provider:
- **Overlay Network**
 - A network built on the top of physical network
 - Nodes are connected by logical links
 - Flat system architecture



Features (Cont.)

- Large-scale resources
 - Heterogeneous
 - Millions of nodes
- Collaboration
 - Based on voluntary participation
 - Global reach
- Flexible, resilient to attacks, anonymous
- ...

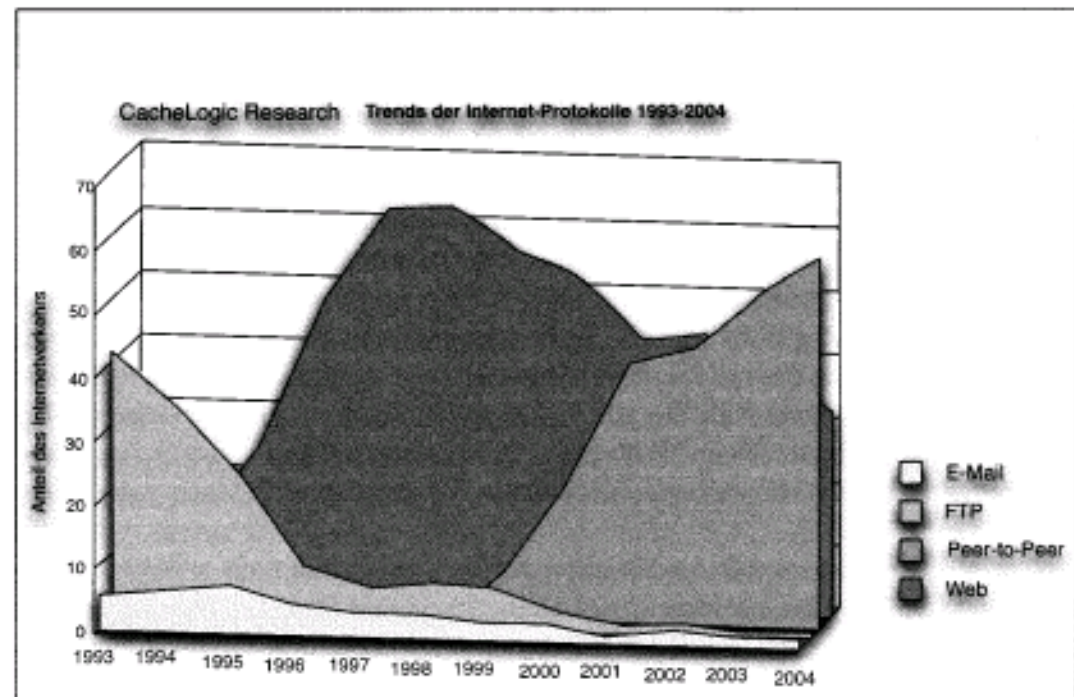
- Please name some typical P2P systems.
- What do you think about P2P applications?

Why?

- Centralized -> decentralized
- Client to Server -> Peer to Peer
- Benefits?

Increasing Requirement of Resource Sharing

- Text, email: 100 KB
- Picture: 1MB
- Music: 10MB
- Movie: 1GB
- HD Video: 10GB+



Developing of Network Bandwidth

- Maximum bandwidth of common Internet access technologies [Wiki]
- Current network bandwidth cannot satisfy user demand
- It needs a more efficient way to share resource

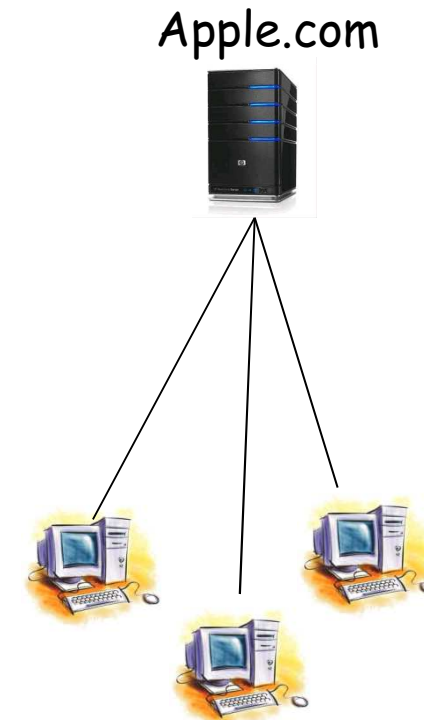
56 kbit/s	Modem / Dialup
1.5 Mbit/s	ADSL Lite
1.544 Mbit/s	T1/DS1
10 Mbit/s	Ethernet
11 Mbit/s	Wireless 802.11b
44.736 Mbit/s	T3/DS3
54 Mbit/s	Wireless 802.11g
100 Mbit/s	Fast Ethernet
155 Mbit/s	OC3
600 Mbit/s	Wireless 802.11n
622 Mbit/s	OC12
1 Gbit/s	Gigabit Ethernet
2.5 Gbit/s	OC48
9.6 Gbit/s	OC192
10 Gbit/s	10 Gigabit Ethernet
100 Gbit/s	100 Gigabit Ethernet

Client/Server

- The client arrives and **requests** a service at any given point in time
- The server is dedicated to the service and **responds** to the client

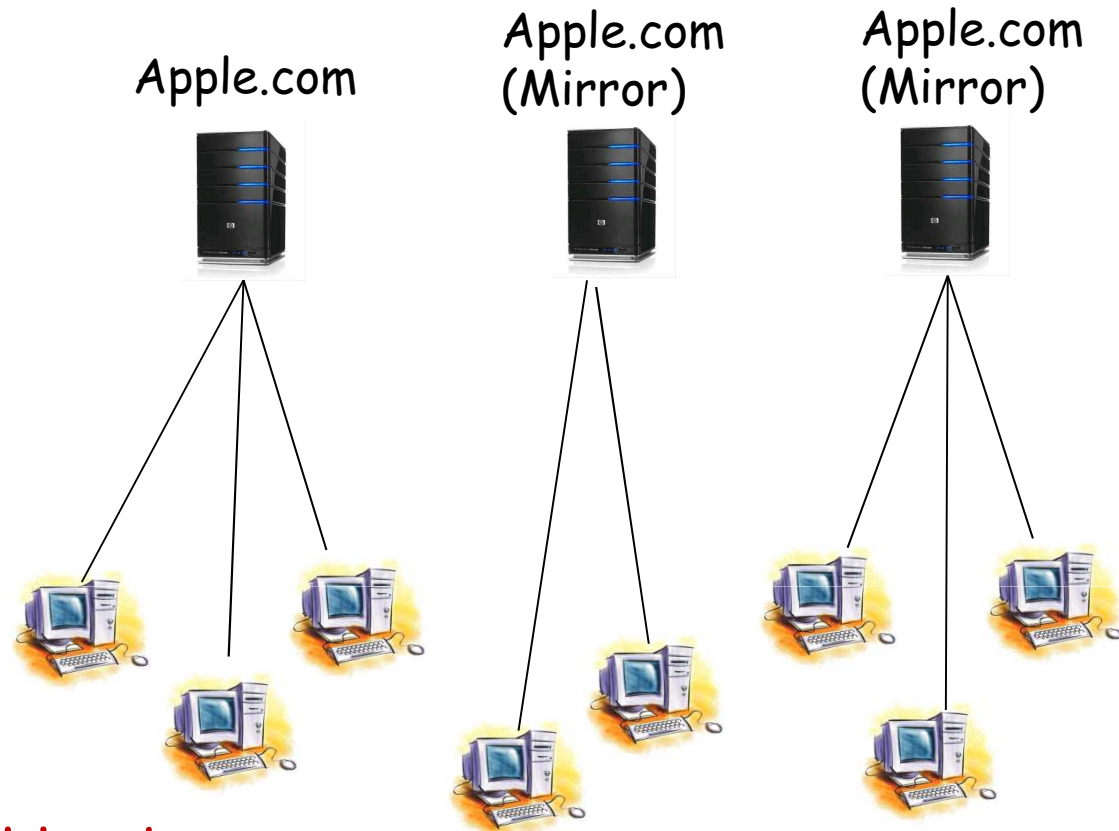
Problems

- Hot spot-uneven workload
- Bottleneck: bandwidth, CPU, ...
- Single point of failure
- Scalability
- ~~Maintenance~~



Replication

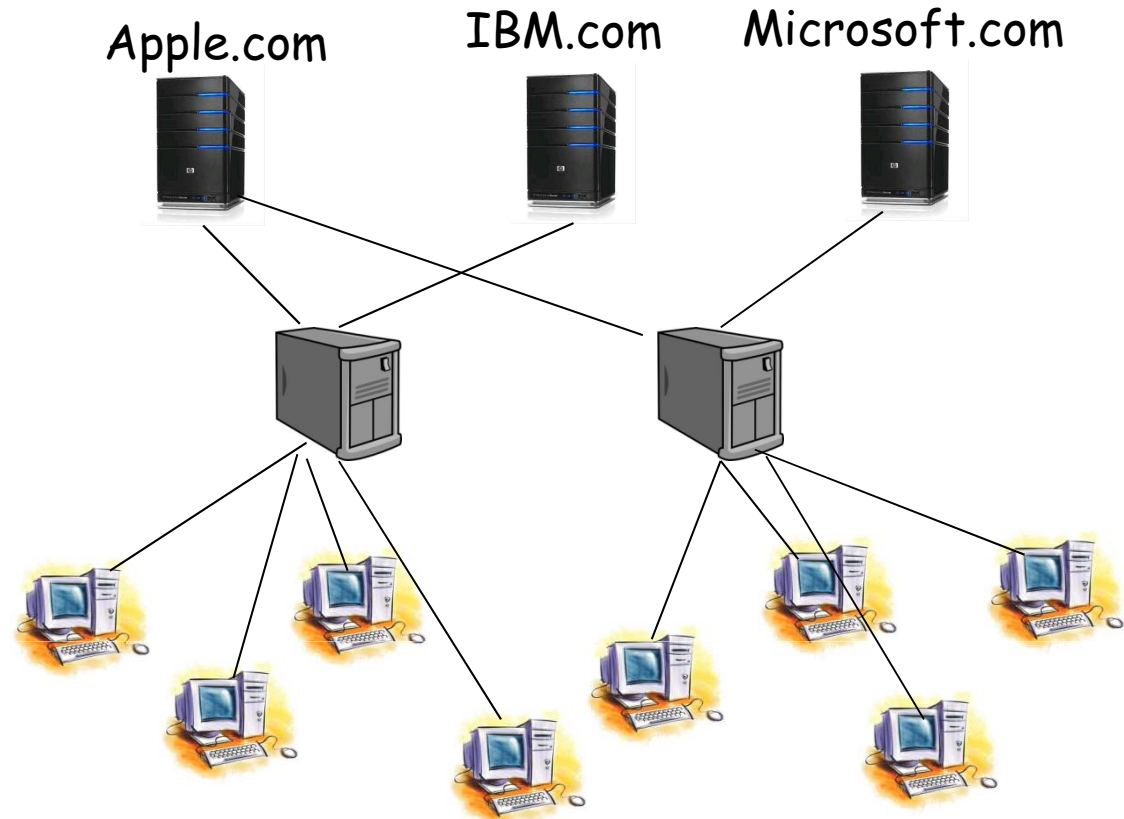
- Replication



Problems

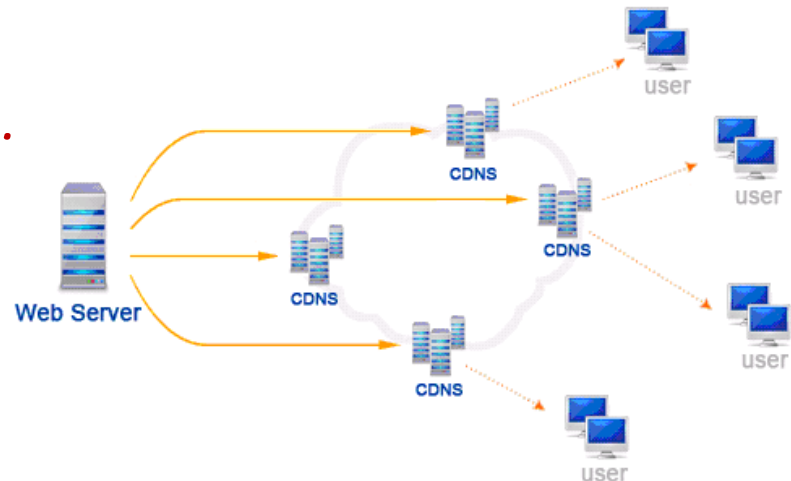
- Hot spot-uneven workload
- Bottleneck: bandwidth, CPU, ...
- ~~Single point of failure~~
- ~~Scalability~~
- Maintenance

Proxy, CDN



Problems

- ~~Hot spot uneven workload~~
- Bottleneck: bandwidth, CPU, ...
- ~~Single point of failure~~
- ~~Scalability~~
- Maintenance



P2P: Advantages

- Changing the way network bandwidth usage
- Easy to deploy, easy to use
- Dynamic for joining and leaving
- Distributed resource sharing
 - Files, data, storage, computation, ...
 - Provide something useful and free
 - Anyone can contribute
- Fault tolerant
- Service ability: large scale
- Service of quality: the more user, the better

How?

- How well is P2P network doing?
- State of art?

P2P Applications and Systems

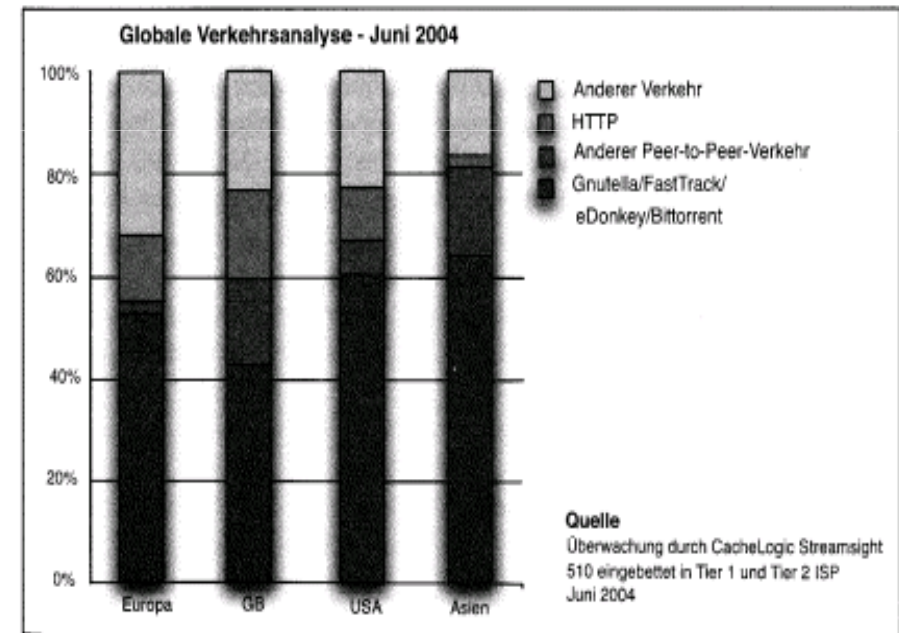
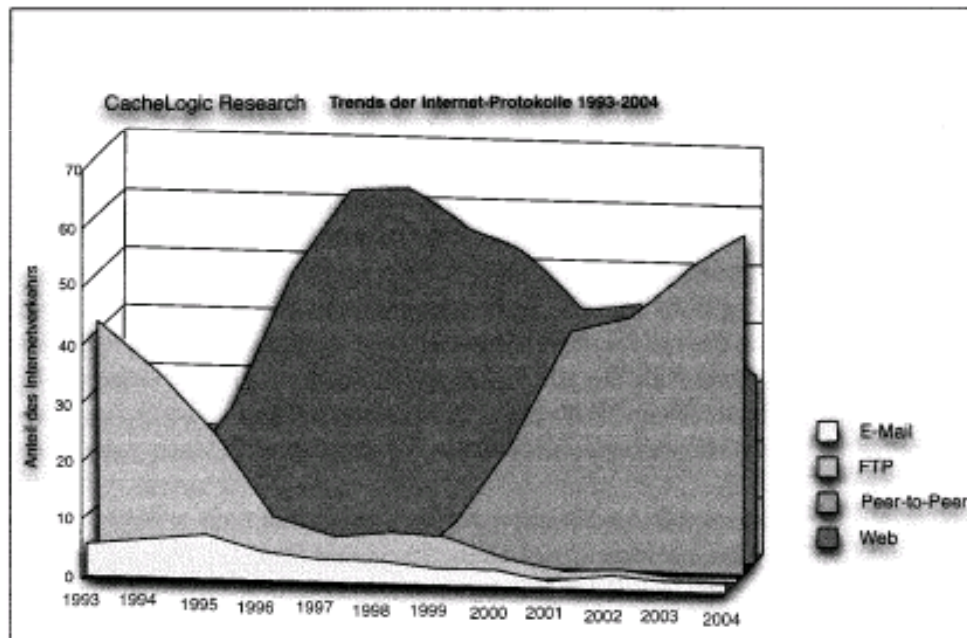
- File sharing
 - Napster, Gnutella, BitTorrent
 - Edonkey, Emule, ...
- Multimedia streaming
 - P2P TV: PeerCast, PPLive, PPStream, TVUnetworks, TvAnts, Zattoo,
 - P2P based VOD systems
- Communication
 - Skype, ...
 - MSN, ...
- Computation
 - SETI@home: Search for Extra-Terrestrial Intelligence

Current State of P2P System

- P2P applications are popular over the world
- P2P networks are mainly used for resource sharing
 - Music, videos, software, ...
 - Some are illegal copyrighted materials
- New emerging applications
 - Online media streaming, P2P TV
 - P2P telephone system
 - Software installation and update
 - Decentralized social network applications

P2P Changes Network Traffic

- P2P applications accounts over 70% of network traffic



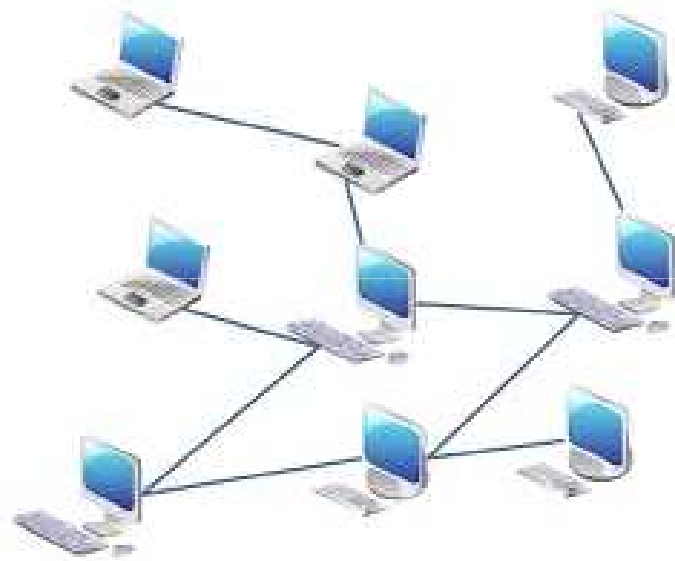
Typical Research Topics

- Structure
 - How to search information
 - Unstructured P2P
 - Structured P2P
- Security and privacy
 - How to protect system security and user privacy?
 - Security in P2P Networks, Anonymity
- Legal issues

Unstructured P2P Networks

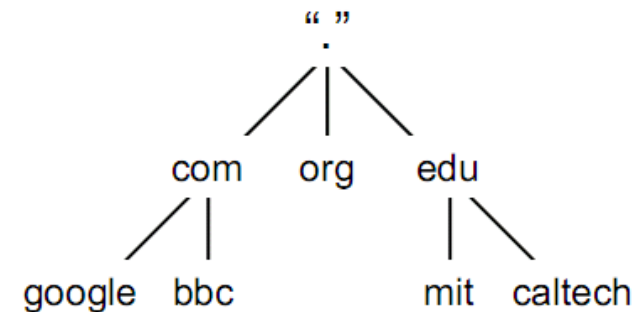
Search in P2P Networks

- How to locate resource in P2P networks?



DNS- Domain Name System

- A hierarchical distributed naming system for computers, services, or any resource connected to the Internet or a private network[Wiki]
- Translates queries for **domain names** into **IP addresses** for the purpose of locating computer services and devices worldwide
- **Distributed database** organized in **hierarchy structure**



Directory

Make Y! your homepage

Web Images Video Local Apps More ▾

YAHOO!

Search

HOME Thu, May 3, 2012

YAHOO! Deutschland

SIGN IN New here? Sign Up

MAIL Check email

YAHOO! SITES

- Autos
- Dating
- Finance (Dow ↕)
- Flickr
- Games
- Horoscopes
- Jobs
- Mail
- Messenger
- Movies
- My Yahoo!
- News
- omg!
- Real Estate
- Screen
- Shine
- Shopping
- Sports
- Travel
- TV
- Weather (52°F)



Gingrich drops out, will support Romney

The former House speaker gives a tepid endorsement to his GOP rival after a turbulent campaign. **His next move** >>

- Obama ad stars Gingrich
- Still \$4 million in debt
- Gingrich campaign spoof

Ex-NFL star Seau is dead Obama's ex-girlfriend Star's sultry Lovelace poster New twist in tourist's death Gingrich ends campaign

1 - 5 of 65

NEWS

Have student loan debt?

If you're one of the estimated 37 million Americans in the red, we want to hear your story.

- Sen. Brown criticized for keeping daughter on health plan
- Trayvon Martin shooter called Mexicans 'wanna be thugs'
- Water guns banned, handguns allowed at GOP convention
- 13 charged in hazing death of FAMU band member
- MIT, Harvard link up with free online courses
- News Corp. board: 'We have full confidence' in Murdoch's 'fitness'

TRENDING NOW

01 Junior Seau dead	06 Alyson Hannigan
02 Patricia Medina dies	07 World's largest McD...
03 Paulina Gretzky	08 Jaleel White
04 Sarah Phillips ESPN	09 Mexico yacht crash
05 Retirement income	10 Student loans

AdChoices



FULL COUNT ON YAHOO! SPORTS

A WHOLE NEW WAY TO WATCH BASEBALL

Check It Out

Ad Feedback

MUST-SEE VIDEOS ON YAHOO!



Information Retrieval System

- Keyword-based



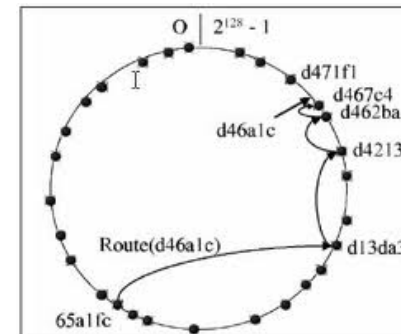
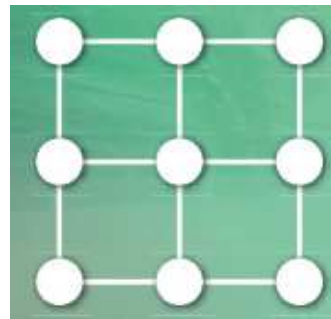
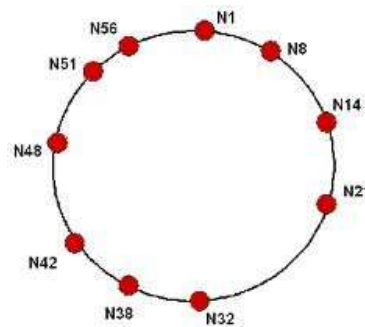
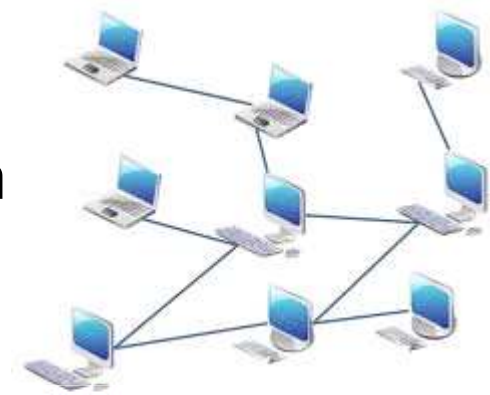
Google Search

I'm Feeling Lucky

Google.de offered in: [Deutsch](#)

Search in P2P Networks?

- Unstructured P2P
 - Highly flexible, dynamic, easy to maintain
 - Hard to find information
- Structured P2P
 - Hard to maintain its structure
 - Easy to find information



Unstructured P2P Networks

- Napster
- Gnutella
- BitTorrent

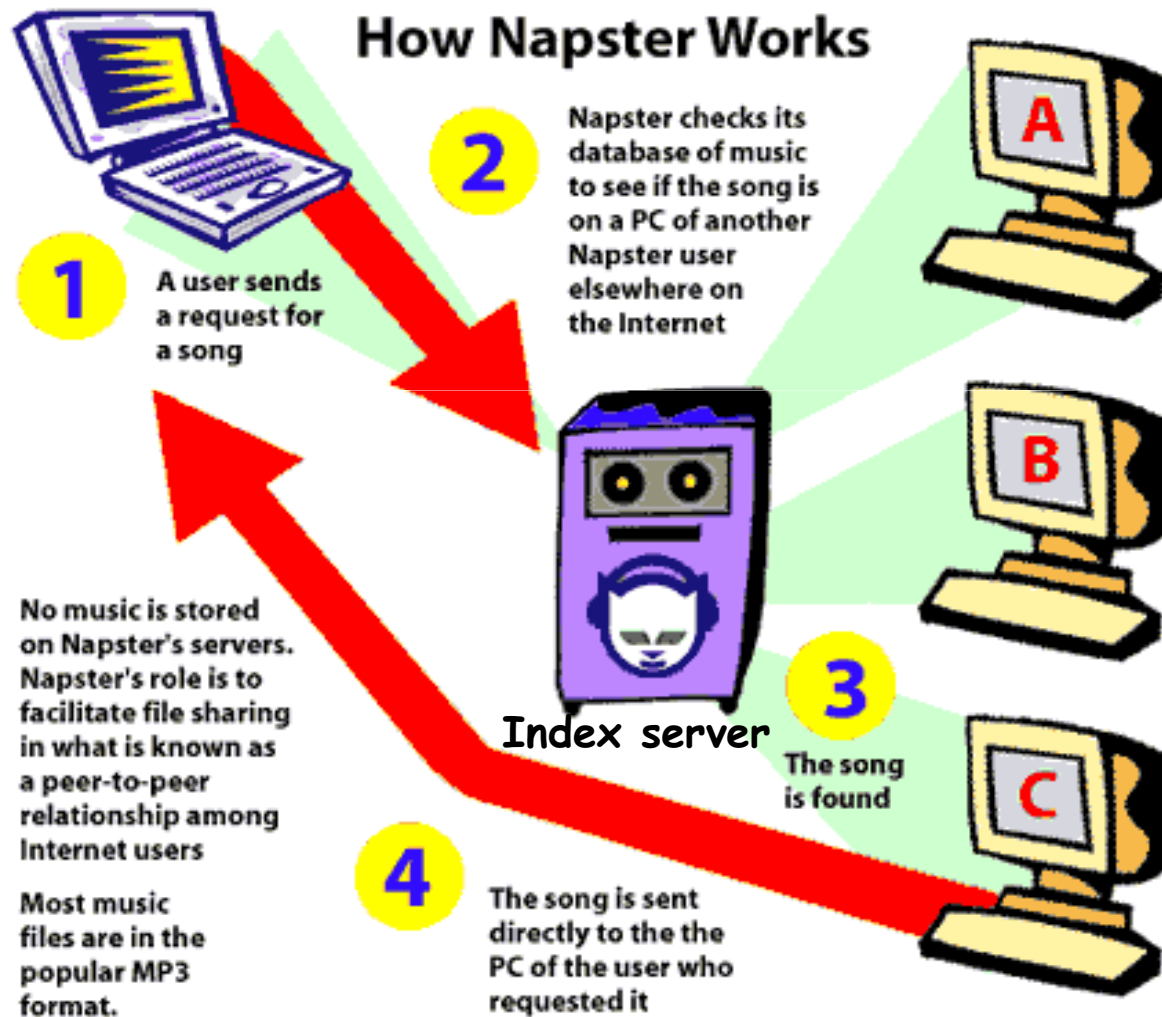
Napster

- The first P2P file sharing applications (1999)
- Mostly sharing MP3 files
- Grew in popularity at 13.6 million users in 2001
- Was shut down in 2001 due to lawsuits
- In 2003 it came back as an online music system (not a P2P anymore)

Search in Napster

- Searching information based on a global index server
 - Step 1: registration
 - User ID, IP address
 - A list of files to be shared
 - Step 2: searching
 - Sending queries containing keywords to the index server
 - Search results were a list of relevant filenames and descriptions
 - Step 3: file download
 - According to the searching results, the node connect to the target peer to fetch the file.
 - Without server involved

An Example



Napster: Pros and Cons

- Strengths
 - Consistent view of the network
 - The central server knows everything
 - Easy to manage
 - Fast and efficient searching
- Weaknesses
 - Single point of failure
 - Bandwidth and CPU bottleneck
 - Hard to scale
 - Central server need to be powerful

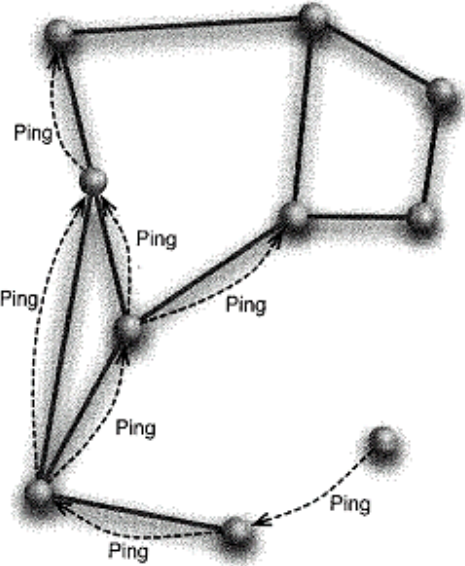
Gnutella 0.4

- Answer to some of Napster's weaknesses
- Pure P2P, fully distributed
- It is a protocol, but cannot be used in a big network
- Currently Gnutella 0.4 is not in use anymore
- Based on overlay network
- Operations
 - Join
 - Search
 - Download: like Napster, occurs directly between peers

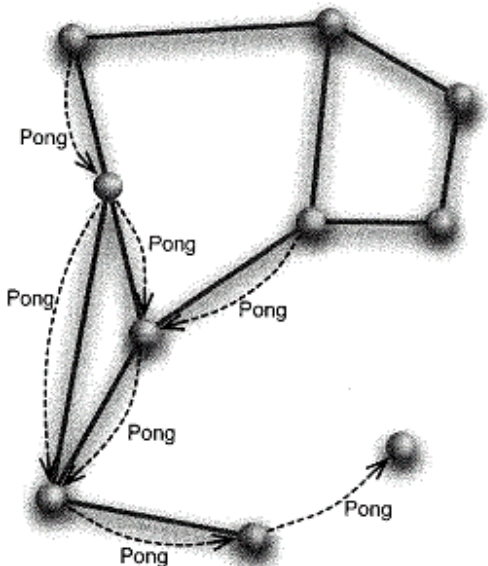
Join

- When a new peer enter the network, it probes the overlay by broadcasting **ping** and a TTL value
- A peer receiving a ping responds with a **pong**, together with its IP address
- Pongs returned along reverse path
- When a pong is received, new links are added between peers

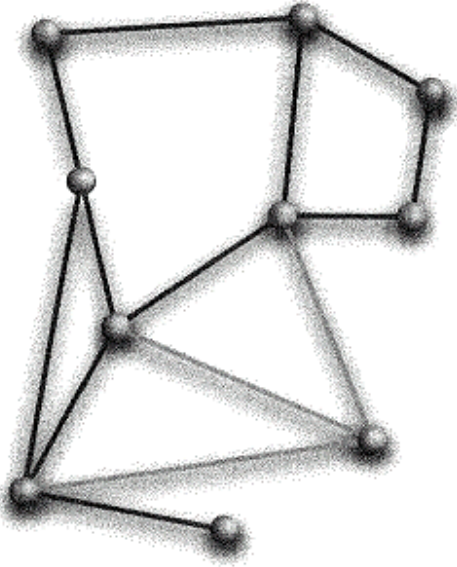
An Example



(a)



(b)

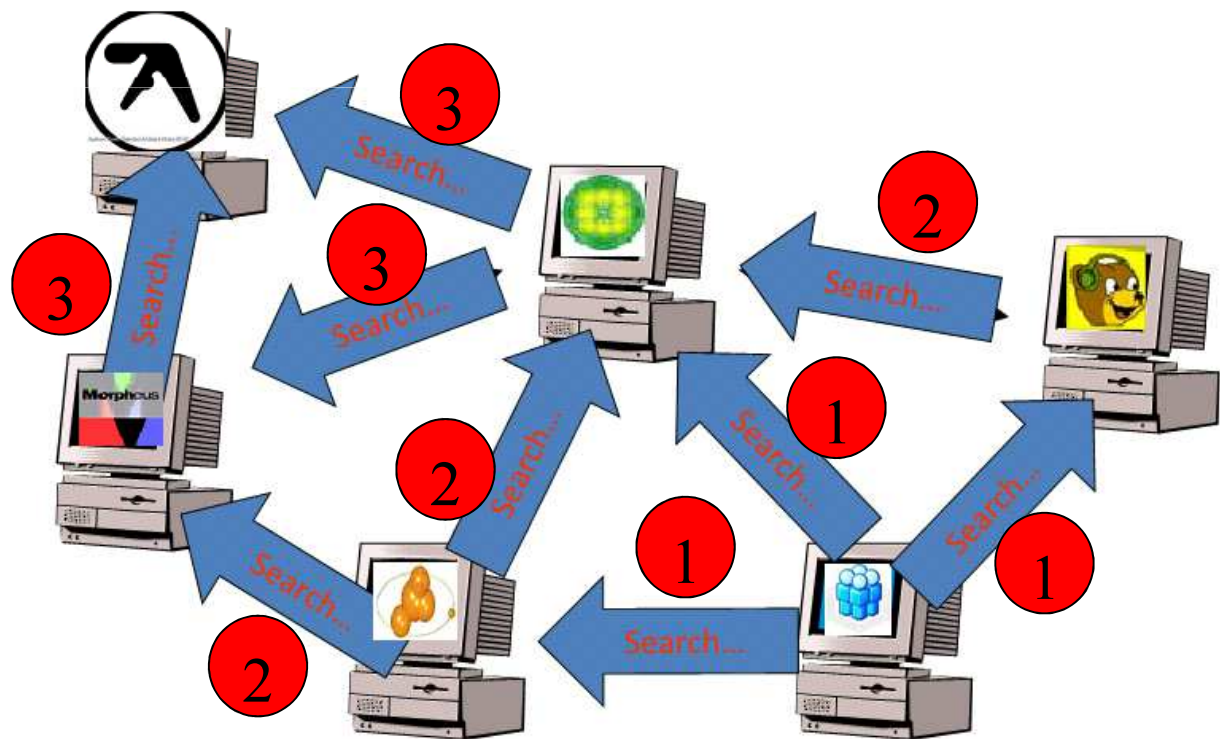


(c)

Search

- Queries are flooded to the whole network
- Peers with the desired files will answer the query
- An example

1



Gnutella: Pros and Cons

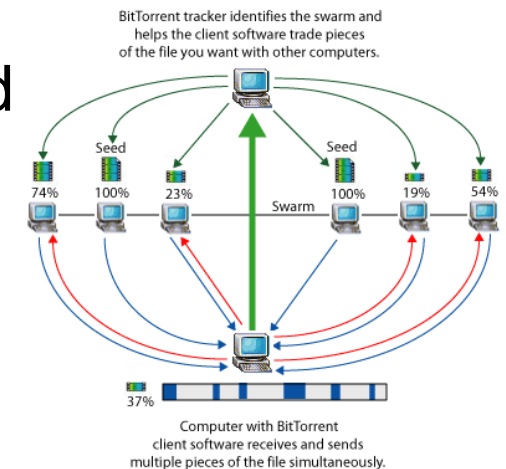
- Strengths
 - Fully distributed
 - Highly fault-tolerant, no single point failure
 - Easy to deploy and implement
- Weaknesses
 - Flooding is highly wasteful of bandwidth and other resources (flooding storm)
 - Solution: limit query radius
 - Maintenance is expensive (ping/pong)

BitTorrent

- A new popular approach to sharing large files
 - It accounts for 30-50% of all Internet traffic
- Originally used for distributing legal content
 - Linux distributions, software updates
 - Official movies
 - Games, ...
- Goal:
 - Quickly and reliably replicate one file to a large number of clients
- Call it “P2P content distribution”

Basic Idea

- Chucking:
 - Files split into smaller pieces or chunks
 - Chunks can be downloaded in parallel
 - Downloading order does not matter
- Swarming
 - Clients join a crowd of peers uploading and downloading the same content
 - Nodes request chunks from neighbors and download content in parallel
- Use the web server to publish content
- Use a central unit to locate resource



Basic Components

- Web server: for content publication
- **Tracker**: a special central server for running the content distribution system
 - Tracking active peers
 - Mapping from file name to peers
- Peer
 - **Seed**: a peer with a complete copy of the file
 - Leecher: peer still downloading the file
- **“.torrent” file**: metadata and description of the file
 - The number of chunks
 - The tracker’s IP



Torrent-file

Tracker: 127.0.0.1
Chunks: 42
Chunk 1: 12345678
Chunk 2: 90ABCDEF

Operation



○ Sharing a file:

- (1) Seed generates a “.torrent” file from the file
- (2) Upload the “.torrent” file to some public web server or sending it to friends by email

○ Searching a file:

- No dedicate search component
- User can search “.torrent” file from web server

○ Downloading a file:

- (1) Download the “.torrent” file
- (2) Connect to the tracker to locate the file
- (3) Choose some fast peers to download chunks in parallel

Tit-for-Tat Policy and Chunk Selection

- Tit-for-Tat policy
 - The more you give, the more you get
 - A peer serves peers that serve it
 - Encourages cooperation, discourage free-riding
- Chunk selection
 - Peers uses **rarest first** policy when downloading chunks
 - Having a rare chunk makes peer attractive to others
 - The goal is to maximize availability of each chunk

BitTorrent : Pros and Cons

- Strengths
 - Works well for “hot content”, very fast and resilient
 - Proficient in utilizing partially downloaded files
 - Discourages “free-riding”
 - Efficient for distributing large files to a large number of clients
- Weaknesses
 - Assume all interested peers active at same time
 - Tracker could be single point of failure
 - Vulnerable to attack
 - Lack of search feature

Next Week

- Structured P2P networks

References

- [1] Stephanos Androutsellis-Theotokis and Diomidis Spinellis. A survey of peer-to-peer content distribution technologies. ACM Comput. Surv. 36(4), 335-371. 2004.
- [2] Hari Balakrishnan, M. Frans Kaashoek, David Karger, Robert Morris, and Ion Stoica. Looking up data in p2p systems. Comm. ACM 46,2(Feb.), 43–48. 2003.
- [3] Clip2: The Gnutella protocol specification v0.4, 2001
- [4] Pouwelse, Johan; et al. "The Bittorrent P2P File-Sharing System: Measurements and Analysis". Peer-to-Peer Systems IV. Berlin: Springer. pp. 205–216. 2005.
- [5] Erik Nygren, Ramesh K. Sitaraman, and Jennifer Sun, The Akamai Network: A Platform for High-Performance Internet Applications, ACM SIGOPS Operating Systems Review, Vol. 44, No.3, July 2010.