

CSC 402 – Internet Technology

Recap

- P2P Model
- P2P – Pros and Cons
- P2P – Types
- Client\Server Model vs. P2P Model
- Taxonomy

Evolution of P2P

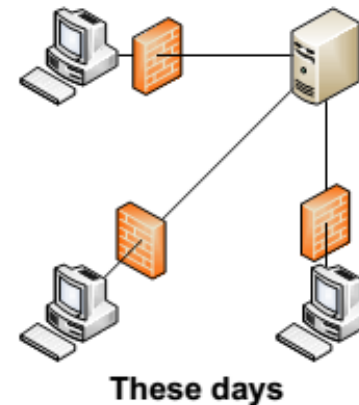
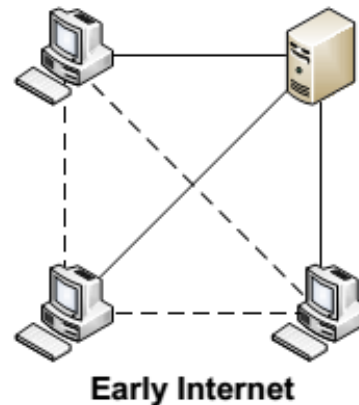
- ARPANET
 - The goal of the original ARPANET : share computing resources around the USA
 - Initially, the ARPANET: 4 independent hosts with equal status:
 - Stanford Research Institute (SRI)
 - University of Utah
 - University of California, Los Angeles (UCLA)
 - University of California, Santa Barbara (UCSB)
- Observations from the early years:
 - The early Internet client/server applications
 - telnet: log-in and use resources of a remote compute server
 - FTP: send and receive files from a file server
- However, notice the following
 - the whole usage pattern is symmetric
 - every host could connect to any other host
 - servers acted as clients as well
- What's important to remember:
- the **Internet started as a P2P system** later on, it has become more restricted to client/server

WWW in a nutshell

- Historic notes:
 - 06.08.91, summary of the WWW project was posted on Usenet. Considered as the debut of the WWW.
 - www.groups.google.com/group/alt.hypertext/msg/395f282a67a1916c
- How it happened?
 - The driving force: disseminate information easily for everyone.
 - Result: a wave of ordinary people began to use the Internet.
 - **NOTE:** WWW is based on the client/server model using HTTP.
- WWW technological implications
 - asymmetric access.
 - Firewalls.
 - network address translators (partially).
 - dynamic IP (partially).
- Bandwidth asymmetry
 - more downlink (server to client) than uplink (client to server)
 - affected access technologies. Example: ADSL2+ (ITU G.992.5, 2003): downstream rate = 24 Mbit/s, upstream rate = 1 Mbit/s
- Asymmetric access: places severe limitations
 - symmetric service became almost "prohibited"
 - bidirectional traffic: TCP performance degradation
 - for more information, see RFC 3449

WWW in a nutshell

- The early days of the Internet:
 - Any host can access any other host. No harm done: pro were using them.
- WWW made the Internet a public place:
 - The need to secure the network protecting hosts from unlimited access. Firewalls is a tool to control access to subnetworks networks and hosts.
- What Firewalls do:
 - Stand in-between the internal network and the Internet outside and choose which traffic to let through and which to deny.

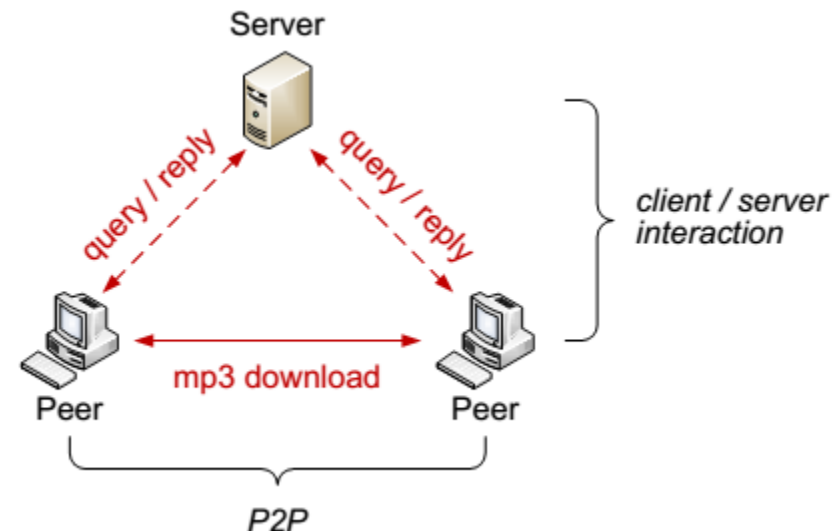


WWW in a nutshell

- A typical firewall do the following:
 - Allows outgoing connections to any external host.
 - Prevents incoming connections from some/all external hosts.
- A host protected in this way cannot easily function as server
 - Outgoing connections may be restricted to certain applications.
 - We block traffic to certain ports at the firewall.
 - e.g., only HTTP (ports 80 and 8080) and FTP (ports 20 and 21).
- IP address assignment: in the early days, static assignment
 - After WWW (DHCP): dynamic due to too many users.
 - **Result:** many hosts on the Internet are not easily reachable.
- Usage of Network Address Translation (NAT)
 - Allows the use private IP addresses within a local network.
 - **Result:** not only the host address is dynamic, it is not even visible!
 - **Consequence:** huge problems for something different than WWW.

Napster

- Some important notes:
 - the driving force: frustration.
 - digital music files were not easy to find back then.
 - no intentions to make history.
 - creative way of thinking in WWW era.
- Developer: Shawn Fanning, 18-year-old student, 1999.
- Functionality: share mp3 files.
- A year later, in 2000, Napster had 20 million users.
- Napster is a hybrid P2P system:
 - peers (Napster clients) and a central server
 - searching via central index containing location of all shared files
 - the actual file transfer occurs between peers



Napster

- Napster worked as follows:
 - After log-in: list of shared files sent to Napster server.
 - Server maintains an index of files of users who are currently on-line.
 - Index is automatically updated when users log-on and log-off.
 - Searching for content.
 - A client sends the query to the server.
 - Query: pattern for search in filenames.
 - Reply: names of files, IP addresses and port numbers.
 - A user decides from which peer to download the files.
 - File download is performed directly between the peers.
- Napster focused exclusively on mp3-encoded music files.
- Only one client implementation, called Napster in 1999.
 - Windows only for one year.
 - In 2000 Black Hole Media: Macintosh client called Macster later used by Napster as the Napster client for Mac.
- The Napster protocol is a **closed-source** protocol
 - i.e., no one knows for sure how searching and transfer is done.

Napster

- The reason for Napster's failure was not technical, but legal:
 - 20 millions of user in 2000
 - Recording Industry Association of America (RIAA)
 - ban the exchange of copyrighted material and/or
 - shut down the network
 - Napster deployed various filters in 2000
 - Napster was forced to shut down in 2001
 - 2002: Napster's brand and logos were purchased by Roxio
 - online music service as Napster 2.0
 - 2008: Napster, Inc. was purchased by Best Buy

Bit Torrent

- BitTorrent was developed by Bram Cohen in 2001
 - Nowadays, it is maintained by Cohen's company – BitTorrent, Inc.
- The BitTorrent network is huge
- E.g., according to isohunt.com (September 15, 2010):
 - Peers: 23.53M (+3.1M over the last 5 months)
 - Shared files: 134.05M (+15.68M)
 - The total amount of shared content: 11,322.96 TB (+715.70 TB)
- Unlike other P2P networks, BitTorrent does not offer the capability to search for content within the application
- Instead, users must have prior knowledge of tracker sites and know where to look for the torrents they want to download
- BitTorrent tracker – a server which assists in the communication between peers using the BitTorrent protocol (Tit-for-Tat Protocol)

BitTorrent

- How BitTorrent works:
 - A user contacts a tracker to find and download a torrent file for the data file he wants.
 - Using this torrent file, the BitTorrent client software communicates with the tracker to find other peers running BitTorrent that have the complete file (aka seeds) and those with a portion of the file (i.e., peers that are usually in the process of downloading the file).
 - The tracker identifies the swarm, which is the connected peers that have the complete file or a portion of it and are in the process of sending and/or receiving it.
 - The tracker helps the client software trade pieces of the file with other peers in the swarm.
 - BitTorrent uses the SHA-1 hash function to determine which pieces of the file are good and which are bad.
 - If the user continues to run the BitTorrent client software after the download is complete, this peer becomes a seed.

BitTorrent

- **Public tracker sites**, such as The Pirate Bay, allow users to search in and download from their collection of torrent files; users can typically also upload torrent files for content they wish to distribute.
- **Private tracker sites**, such as Demonoid, operate like public ones except that they restrict access to registered users and keep track of the amount of data each user uploads and downloads, in an attempt to reduce **leeching**.
 - Users who have low upload ratios may see slower download speeds until they upload more.
 - Some trackers exempt dial-up users from this policy, because their uploading capabilities are limited.

BitTorrent

- **Seeder** is someone from whom you can download a piece of file. Hence they affect the overall availability of file on P2P network.
- **Leecher** is someone who has downloaded a file but is not sharing it back to P2P network. Hence, overall availability of file decreases.
- **Peer** is someone who is involved in file sharing activity. It is a generic term.
- Users who share no files referred to as **freeloaders** (or free riders).
- Importantly, leechers are not necessarily free-riders.
- Altruistic seeders help bootstrap (self-starting process without external input) new peers and provide a significant fraction of the global upload bandwidth.