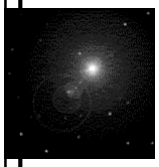


A Course on Foundations of

Peer-to-Peer Systems & Applications

CS 6/75995
Foundation of Peer-to-Peer

Applications & Systems

Kent State University

Dept. of Computer Science

www.cs.kent.edu/~javed/class-P2P08/

Today's Topic

Unit background and administrivia

Foundations of Peer-to-Peer Applications & Systems

INTERNET APPLICATION & SYSTEM DESIGN

Peer to Peer System
Past & Present

Overview

1. Status Quo: Networks (Over)Filled with Peer-to-Peer Traffic
2. How It All Began: From Arpanet to Peer-to-Peer
3. The Napster Story
4. Gnutella and its Relatives: Fully Decentralized Architectures
5. Driving Forces Behind Peer-to-Peer


1) Freenet
2) Buzzpad
3) WuWu

INTERNET APPLICATION & SYSTEM DESIGN

What is P2P

P2P systems are overlay architectures, with the following characteristics:

- Two logically separate networks
- Mostly IP based
- Decentralized and self organizing
- Employ distributed shared resources (computing power and data storage)
- Initially developed for file-sharing
- Various realizations
- Common basis for signaling: IP (TCP and UDP)
- Common basis for data transmission: HTTP
- Use flooding in the overlay to a certain extent

INTERNET APPLICATION & SYSTEM DESIGN

Definition?

- Oram:
- *[a peer to peer system is] a self organizing system of equal, autonomous entities (peers) [which] aims for shared usage of distributed resources in a networked environment avoiding central services.*

INTERNET APPLICATION & SYSTEM DESIGN

Decentralized Resource Usage

- Resources (bandwidth, storage, processing power) are used in a manner equally distributed as possible and are located at the edges of the network, close to the peers.
- With a set of peers, each utilized the resources provided by the others. Resources are not only storage or bandwidth but also human presense, connectivity etc.
- Peers are interconnected through a network and distributed globally. Peers however does not need fixed IP.

INTERNET APPLICATION & SYSTEM DESIGN

Decentralized Self-Organization

INTERNET APPLICATION & SYSTEM DESIGN

- Peers interact directly with each other and decide resource sharing. Does not require a centralized authority for co-ordination.
- Peers exchange data directly too- without requiring any centralized exchange point.
- Any peer can act both as client and server.
- Peers are equal partners. Each peer is fully autonomous regarding its respective resoruces.

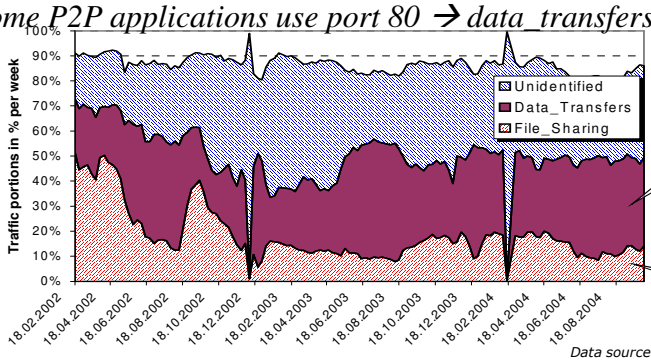
Impacts of P2P

INTERNET APPLICATION & SYSTEM DESIGN

- Rising flow sizes (60 kbyte -> 2 Gbyte)
- 30%-60% of the traffic in the Abilene backbone is caused by P2P applications
- 70% of the traffic in the German Research Network (DFN) is caused by P2P applications.
- T-Online observes an increasing symmetry at the access-level.
- LRZ (Munich Network Center) observes an increasing symmetry between US and Europe

Impacts of P2P at the Abilene Backbone

- ***Unidentified + data_transfers + file_sharing causes 90% of the traffic***
- ***Unidentified traffic and data_transfers increased significantly***
  - *Parts of P2P is hidden (port hopping,…)*
  - *Some P2P applications use port 80 → data_transfers*



Core of Internet2 infrastructure, connecting 190 US universities and research centers

Possible data transfers

Only Signaling

Data source: http://netflow.internet2.edu/weekly/

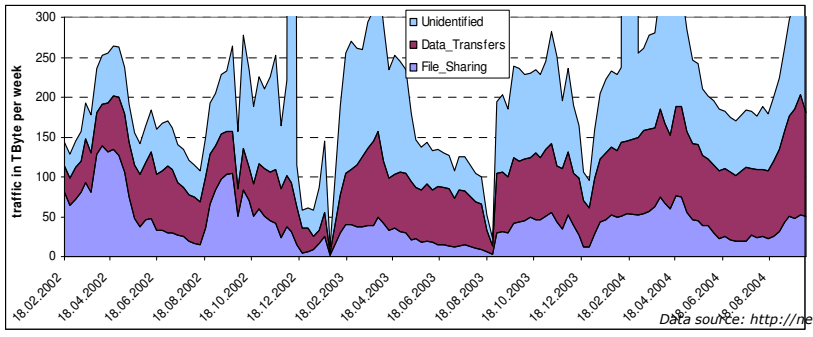
INTERNET APPLICATION & SYSTEM DESIGN

Impacts of P2P at the Abilene Backbone

•P2P Traffic amount (only signaling)
- Is still high (~50 TByte per week)
- Becomes a constant part of the traffic (since end 2002)

•Slumps are assumed to be caused by
- Port closures (firewalls, NATs)
- Verdicts (Napster Case,…)

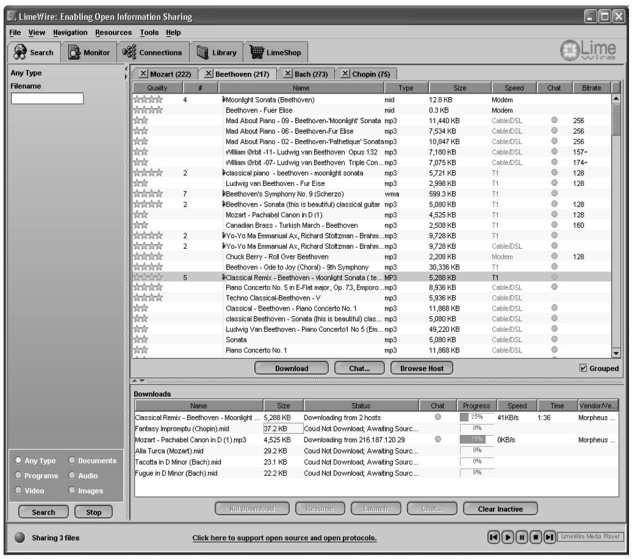
•Data Transfers are caused presumably to a large extent by P2P apps



Data source: http://netflow.internet2.edu/weekly/

INTERNET APPLICATION & SYSTEM DESIGN

Reason for These Experiences



INTERNET APPLICATION & SYSTEM DESIGN

How it Begun: From ARPANET to Peer-to-Peer

- Late 1960s: Establishment of the ARPANET
  - Goal: share computing resources and documents between US research facilities
  - Every host treated equally
  - BUT: the virtual network matched the physical network to a large extent
  - Applications: FTP and TelNet → client server mode, with no decentralized search and storage
  - Central steering committee to organize the network
- 1979: Development of the UseNet protocol
  - Newsgroup application to organize content
  - Self organizing approach to add and remove newsgroup servers
  - Application itself is still a client server application
- ~1990 rush of the general public to join the Internet
  - Applications following the client server approach: WWW, email, streaming
  - Based on modem connections via the SLIP and PPP protocol
  - Straightforward model to administrate and control the content distribution
  - Security concerns resulted in a partitioned Internet by firewalls

INTERNET APPLICATION & SYSTEM DESIGN

The Napster Story

INTERNET APPLICATION & SYSTEM DESIGN

- MAY 1999: Disruption of the Internet community
  First Generation of P2P
  - Introduction of Napster
  - User not only consume and download content but also offer and provide content to other participants
  - Users establish a virtual network, entirely independent from physical network and administrative authorities or restrictions
  - Basis: UDP and TCP connections between the peers
- December 1999: RIAA files a lawsuit against Napster Inc.
  - Target of the RIAA: the central lookup server of Napster
- February 2001: 2.79 billion files exchanged via the Napster network per month
- July 2001: Napster Inc. is convicted
  - Napster has to stop the operation of the Napster server
  - Napster network breaks down
  - BUT: Already a number of promising successors available

Gnutella and Its Rel

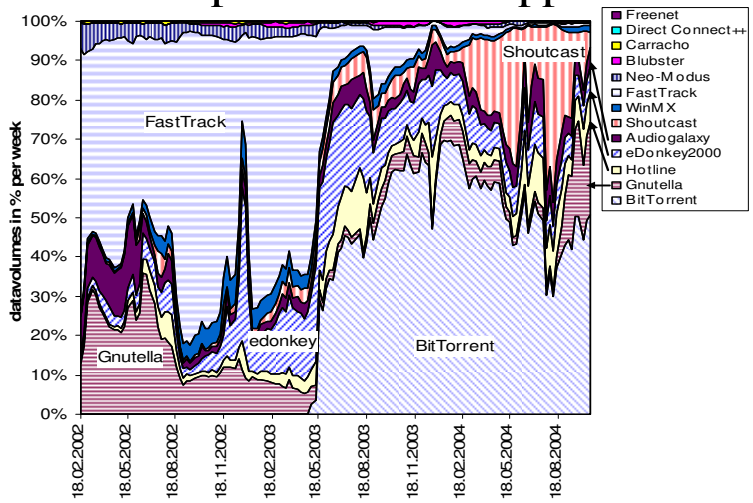
INTERNET APPLICATION & SYSTEM DESIGN

- March 2000: Nullsoft releases Gnutella as an open source project
  - Major developer: Gene Khan
  - Additionally to servent functionality, the peers also take over routing tasks
  - Fully decentralized, no central lookup server → no single point of failure
- October 2000: introduction of hierarchical routing layers. Second Generation of P2P
  - Gnutella: reflector/Superpeer concept
  - Increases the scalability significantly
- Variety of similar fully decentralized P2P-protocols followed soon:
  - Audiogalaxy
  - FastTrack/KaZaA
  - iMesh
  - Freenet

Gnutella and Its Relatives. The story goes on

- August 2001
  - Users adapt very fast to the breakdown of Napster
  - Already 3.05 billion files exchanged per months via the Gnutella network
- Year 2001: Third Generation of P2P initiated
  - First research started on the third generation of P2P, so called structured P2P networks
  - Basic characteristics: Usage of a proactive routing algorithm based on Distributed Hash tables (DHTs)
- August 2002
  - Amount of exchanged data in KaZaA decreases, caused by a high number of defected files (reason: weak hash keys to identify files)
  - Edonkey and Gnutella regain popularity
- May 2003
  - Bittorrent is released
  - Soon causes majority of the observed traffic. Reason: Its popularity, but also that user data is exchanged via the signaling channels in contrast to Gnutella, edonkey,…
- Middle of 2003
  - Beyond the exchange of content, new concepts are developed to use P2P also for other applications
  - Skype a Voice over P2P application is developed
- Today:
  - Major efforts are made to increase the reliability of P2P-searches, to use P2P also in mobile networks, …
  - Ebay buys in the Middle of 2005 Skype to use the paradigm for the communication between the Ebay bidders and sellers

Development of P2P Applications



Traffic portions of the different P2P applications and protocols from the traffic measured per week in the Abilene backbone from 18.02.2002 until 18.010.2004

Data source: http://netflow.internet2.edu/weekly/

INTERNET APPLICATION & SYSTEM DESIGN

Driving Forces Behind Peer-to-Peer

The development of the physical and the technical capabilities of the networks and the participating capabilities

INTERNET APPLICATION & SYSTEM DESIGN

Development of the terminal capabilities:

- 1999:
  - Average hard disk size: ~0.3Gbyte
  - Average processing power (clock frequency) of personal computers: ~ 100MHz
- 2002:
  - Average hard disk size: 100 Gbyte
- 2004:
  - Average processing power (clock frequency) of personal computers: ~ 3GHz
  - → Personal computers have capabilities comparable to servers in the 1990s

Development of the communication networks:

- Early 1990s: private users start to connect to the Internet via 56kbps modem connections
- 1997/1998
  - first broadband connections for residential users become available
  - cable modem with up to 10Mbps
- 1999
  - Introduction of DSL and ADSL connections
  - Data rates of up to 8.5Mbps via common telephone connections become available
  - The deregulation of the telephone market shows first effects with significantly reduced tarriffs, due to increased competition on the last mile
  - → bandwidth is plentiful and cheap!

Next Class:
Early Architectures