HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Peer-to-Peer Systems

Winter semester 2014

Jun.-Prof. Dr.-Ing. Kalman Graffi

Heinrich Heine University Düsseldorf

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Peer-to-Peer Systems

Organizational Information

Organisational matters

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Lecturer

- Jun.-Prof. Dr.-Ing. Kalman Graffi
  - graffi@cs.uni-duesseldorf.de
  - Meetings can be scheduled via emails
- Andreas Disterhöft, M.Sc.

Lecture consists of 3 parts

- Lecture course
- Exercises (included in the lectures)
- 1 Programming project

Announcements / Slides / Recordings

- Information / Slides / Exercises
  - http://tsn.hhu.de/teaching/lectures/2014ws/p2p.html
- Recordings:
  - https://www.youtube.com/playlist?list=PLFJGmP04pevmyZpwVDGmgU4Tq8GQ5-_UG

Organisational matters

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Lecture courses

- Peer-to-Peer Systems
- Time: Thursday, 10:30 – 12:00
- Place: 26.11.6E

Exercises

- Time: Wednesday, 12:30 – 14:00, two weekly
- Starting 22.10.2014
  - 22.10., 05.11, 19.11., 03.12., 17.12. … (see website)
- Place: 25.12.02.33

One practical exercise

- Covers the usage of a p2p simulator
- Overlay evaluations
- Takes the time of two exercises

Lecture

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Goal

- Getting to know the state of the art in p2p networking
- Understand the main design principles
- Get insights in design and evaluation of p2p solutions

Participation

- Presence is optional
- But you miss many details if following only slides & recordings

E-Learning

- All slides and exercises will be on the website
- All recordings will be on Youtube

Testing of p2p software to support the lecture

- WebP2P: Browser-based P2P Audio/Video Chatting
- P2P Social: P2P-based online social network

Topics – preliminary schedule

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

1. Decentralization, Overlay Networks, Characteristics of P2P Systems
2. Unstructured P2P Overlays (Gnutella, edonkey, Kazaa, BubbleStorm)
3. Distributed Indexing and Distributed Hash Tables
4. Structured P2P Overlays (Chord, CAN, Pastry, Kademlia)
5. Interconnection Overlay Networks (Hypercube, De Bruijn Network)
6. Location-based, Heterogeneous Structured P2P Overlays
7. Improvements of P2P Overlays, Scale Free Networks
8. Security Issues in P2P Systems, BitTorrent
9. P2P Streaming, P2P Storage Fundamentals
10. P2P Storage Systems (CFS, Ivy),
11. Monitoring P2P Systems (Sampling, Gossiping, Tree-based)
12. Control Loops in P2P Systems, Monitoring Peer-specific Information
13. P2P-based Social Networks (Safebook, P2PSocial)

Exercise

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Goals:

- Application and deepening of the learned mechanisms
- Seeing and discussion of alternative approaches

Process:

- Exercise are included in the lecture!
- Questions are presented “in line” of the lecture
- Two versions of the slides will be available
  - With questions, no solutions
  - With questions and solutions

In order to register for the exam:

- Register in HIS-LSF

Exam

Applicability

- “Schwerpunkt” or “Wahlpflicht” in Master (Computer Science)
- Field: practical / technical computer science
- Check the applicability in your “Prüfungsordnung”

Date

- Exam will be in February / March

Exam: Written OR oral

- Both topics in the lecture AND in the exercises, on all topics
- Written: 120 minutes OR
- Oral: 30 minutes

Literature

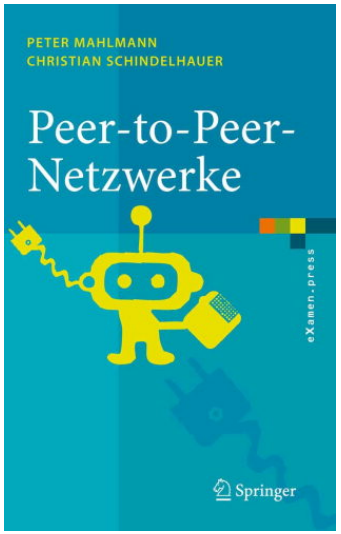
In general
- scholar.google.com
  - „Survey on …“
- Not Wikipedia

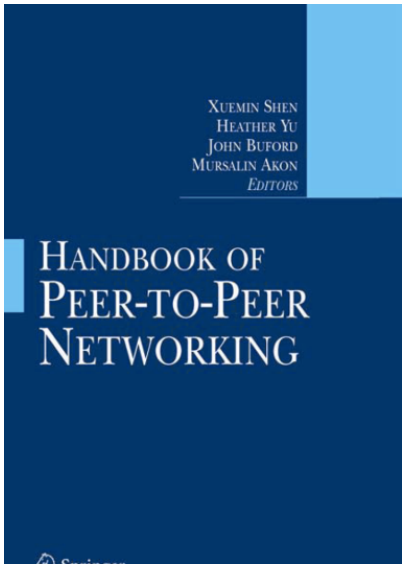
Peer-to-Peer-Netzwerke
- Mahlmann, Schindelhauer

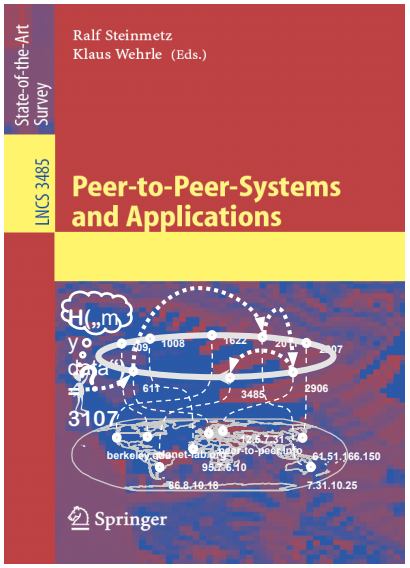
Handbook of P2P Networking
- Shen, Yu, Buford, Akon

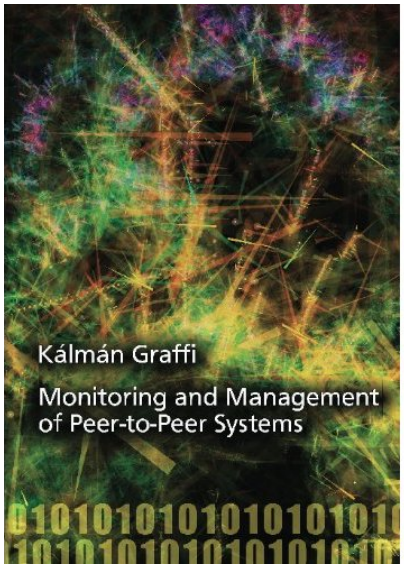
Peer-to-Peer Systems and Applications
- Steinmetz, Wehrle (Editors)
- springerlink.com/content/g6h805426g7t

Monitoring and Management of Peer-to-Peer Systems
- Graffi
- http://tuprints.ulb.tu-darmstadt.de/2248/









HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Technology of Social Networks - An Overview

Jun.-Prof. Dr.-Ing. Kalman Graffi

Overview

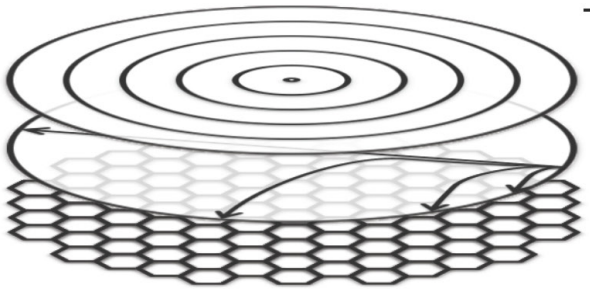
HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Technologies

routing protocols
distributed data storage
in-network processing
security and trust

Networks

self-organizing overlays
peer-to-peer networks
opportunistic networks
vehicular ad hoc networks



Technology of Social Networks

Applications

social networks
on-the-fly computing
emergency response
smart environments

Foundations

performance and costs
quality of service
security and resilience
privacy and anonymity

Networks

Peer-to-Peer Networks

- Network of equal participants -
- Freedom to create logical topologies, harness node resources
- Strong heterogeneity, churn (on/offline behavior), no trust
- Focus on data management

Opportunistic Networks

- Mobile Ad Hoc Networks that span „local communication islands“
- Mobility of nodes
  - → connect islands over time → delay tolerant communication
- Focus on routing / communication

Combination of both

- Mobile decentralized network with p2p network on top

Technologies

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Overlays / routing

- With special focus on social contacts, locality, anonymous routing, robustness against partitioning

Data storage

- Replication and consistency

Publish / subscribe

- And further communication patterns

Multicriterial search

- Search for data or nodes based on various criteria

Monitoring

- Statistics on the network / node capacities / data popularity

Always in combination with security

Applications

Decentralized Online Social Networks

- Motivation: security and robustness
- Goal: decentralized, secure p2p framework for social networks
- Challenges:
  - P2P mechanisms: overlay, routing, data storage, monitoring, …
  - Security: identity management, authentication, access control
  - Application specific requirements: distributed data structures

Online Participation / Normsetting

- Motivation: security and robustness
- Goal: secure and robust platform for online deliberation and voting
- Challenges:
  - P2P / opportunistic networks: routing, data storage, voting
  - Security: Voting restrictions, censorship, blocking of discourse

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Peer-to-Peer Systems - Chapter 1

Decentralization

Overlay Networks

Characteristics of Peer-to-Peer Systems

A few Definitions for Peer-to-Peer Systems

Peer-to-peer systems and applications are distributed systems without any centralized control or hierarchical organization, where the software running at each node is equivalent in functionality. [...] The core operation in peer-to-peer systems is efficient location of data items.

- I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, “Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications”

Peer-to-peer systems can be characterized as distributed systems in which all nodes have identical capabilities and responsibilities and all communication is symmetric.

- A. I. T. Rowstron and P. Druschel, “Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems“

The sheer scale and dynamism in which P2P networks are supposed to operate make the design of P2P systems challenging even for relatively simple applications.

- M. Naor and U. Wieder, “Novel architectures for p2p applications: the continuous-discrete approach“

Detailed Characteristics - Properties

1. Self-organizing system
   - Relevant mechanisms performed by peers
     - No central control
     - Decentralized resource search, allocation and scheduling
   - (Sometimes, servers assist → centralized p2p systems)

2. Combined client and server functionality
   - Resources provided by end systems
     - Storage, communication (forwarding messages)
   - Mostly similar rights – same code!
     - Roles based on capabilities

3. Direct interaction between peers (= “peer to peer”)
   - Provision of services, such as: search, data hosting, communication

Detailed Characteristics - Challenges

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

4. Relevant resources located at (private) nodes (peers)
   - Uncontrolled, voluntary offers
   - Widely spread
   - Often operating behind firewalls or NAT gateways
   - Requires proper mechanism to find and use

5. Capacities of peers are heterogeneous
   - Bandwidth, CPU power, storage space, …
   - Quality depends on device / connectivity

6. Churn: variable connectivity
   - Peers are online for a limited time
   - Very unpredictable, not reliable

Short Question

Is this sufficient to describe p2p systems as we know it?

What other network types do you know with:

- 1. Self-organizing system
- 2. Combined client and server functionality
- 3. Direct interaction between peers (= “peer to peer”)
- 4. Relevant resources located at (private) nodes (peers)
- 5. Capacities of peers are heterogeneous
- 6. Churn: variable connectivity

Others vs. P2P Technology

Mobile ad hoc networks:

- No communication infrastructure available
- Nodes provide bandwidth for common goal
  - to enable communications
  - main issue: routing
- More hop-2-hop than p2p

Sensor networks

- Ad hoc networks with simple devices
- Main challenge: pulling information from sensors

IP Networks

- Interaction of IP routers

Friend-of-a-friend

- Federation of personal webservers
- Linking to trusted friends

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

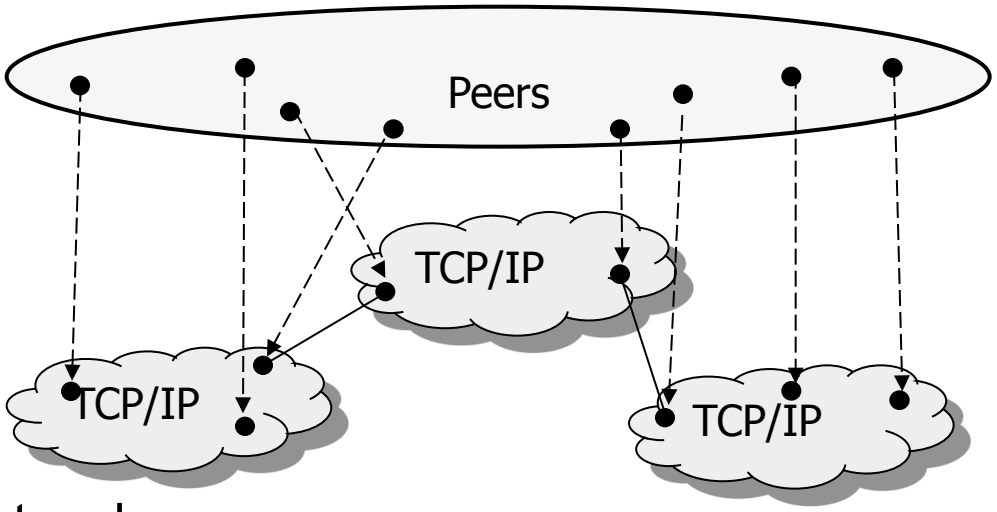
Peer-to-Peer Systems

Definitions and Motivation

– Overlay Networks and P2P Properties

This slide set is based on the lecture "Communication Networks 2" of Prof. Dr.-Ing. Ralf Steinmetz at TU Darmstadt

Overlay Networks



Overlay Network

Underlay Networks

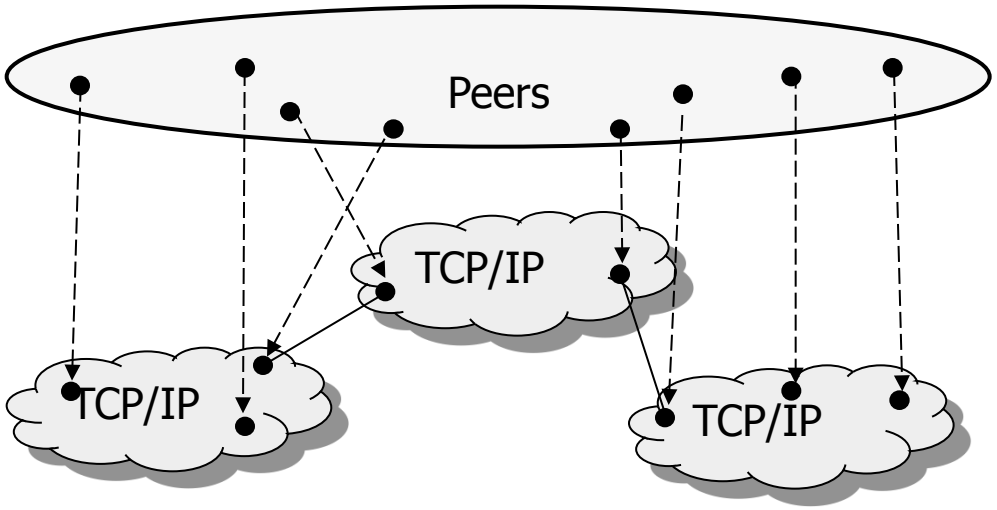
A network

- interconnected nodes
- provides services (service model)
- defines how nodes interact
- needs for addressing, routing, …

Overlay network

- = network built ON TOP of one or more existing networks
- adds an additional layer of
  - abstraction
  - indirection/virtualization

Overlay Networks



Overlay Network

Underlay Networks

E.g. P2P networks form an overlay network

- on top of the Internet (IP network)

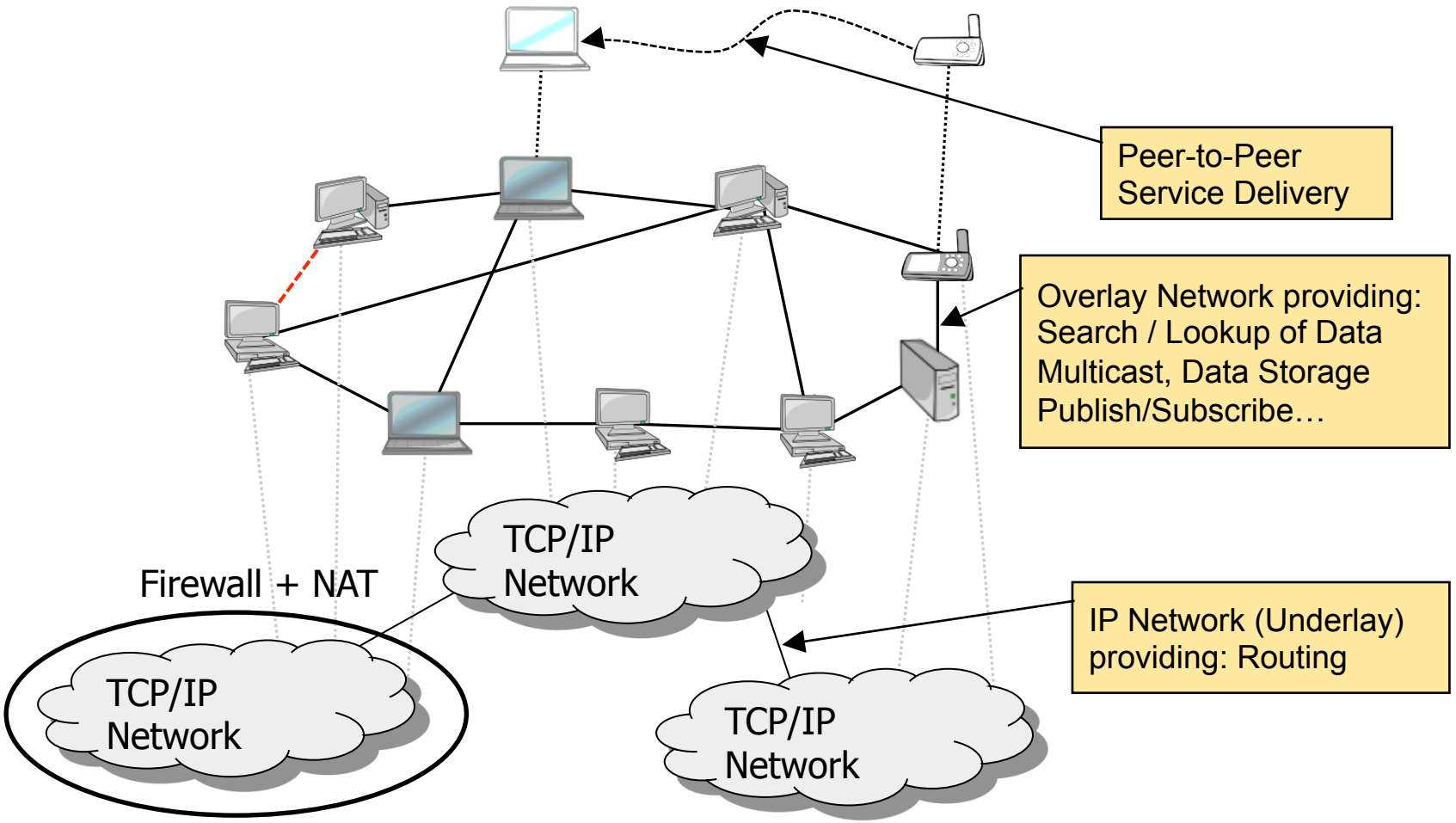
IP networks form an overlay network

- politically and technically
- over the underlying telecom infrastructure

Both introduce new functionality

- IP: routing across networks
- P2P:
  - Search / lookup of data
  - Addressing of users
  - Multicast, Pub/sub …

Schematic View on P2P Systems



Overlay Networks: Advantages

Introduction of a new functionality

New layer fastens search/lookup of requested information

- additional layer solves this problem for higher layers

Do not have to

- deploy new equipment
- modify existing software/protocols

Allow for bootstrapping

- Make use of existing environment by adding new layer

E.g., adding IP on top of Ethernet

- does not require modifying Ethernet protocol or driver

Overlay Networks: Disadvantages

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Overhead

- Adds another layer in networking stack
- Additional packet headers, processing

Complexity

- Layering does not eliminate complexity, it only manages it
- More layers of functionality
  - Introducing interdependecies between layers
- Misleading behavior
  - E.g. corruption drops on wireless links interpreted as congestion drops by TCP

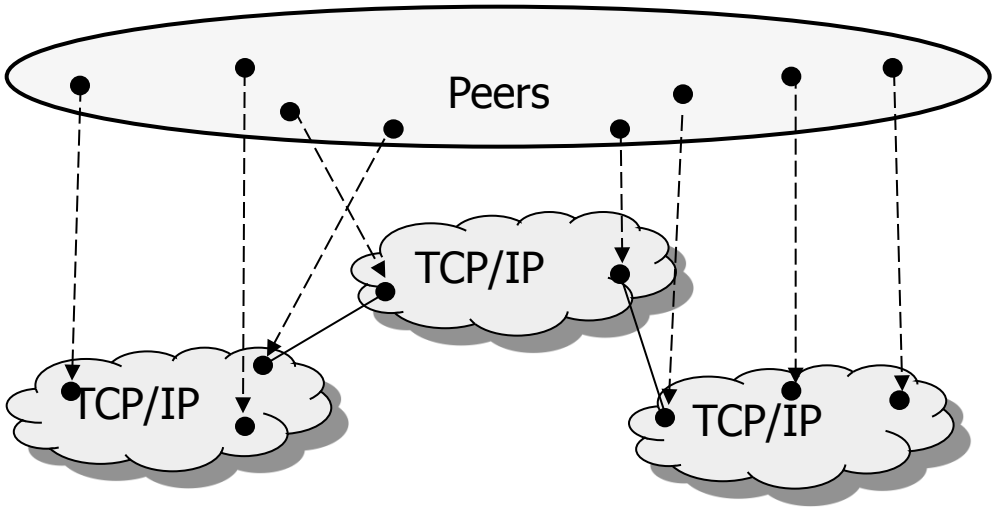
Redundancy

- Features may be available at various layers

Some restricted functionality

- Some features that a “lower layer” does not provide cannot be added on top
  E.g. non real-time capabilities (for QoS)

Overlay Networks: Others



Overlay Network

Underlay Networks

Peer-to-Peer overlay network

- Content-centric networking / routing
- Storage and retrieval
- Search / recommender services
- Friend-of-a-friend network
- User-centric networking

Other (non P2P) overlays

- VPNs (virtual private networks)
- IP over ad hoc networks
- Application-layer multicast
- TOR – anonymizer proxies

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

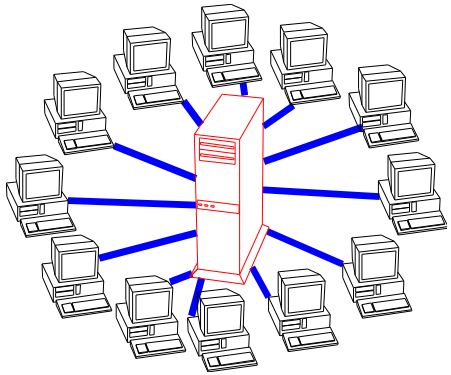
Peer-to-Peer Systems

Definitions and Motivation

– Motivation for Decentralization

This slide set is based on the lecture "Communication Networks 2" of Prof. Dr.-Ing. Ralf Steinmetz at TU Darmstadt

Client / Server Model vs. P2P Technology



Situation:

- 1 server, n * clients

Issue: ???

- E.g. on which server is the information wanted?

Solution:

- Look it up on another server
  (or Google, which does this for you)

Advantages:

- Reliable, well known behavior

Drawbacks:

- Server needs to provide (almost) all resources

Client / Server model is not P2P:

- Communication only between clients and server,
  not between clients and clients

Cloud Computing vs. P2P Technology

Cloud and P2P

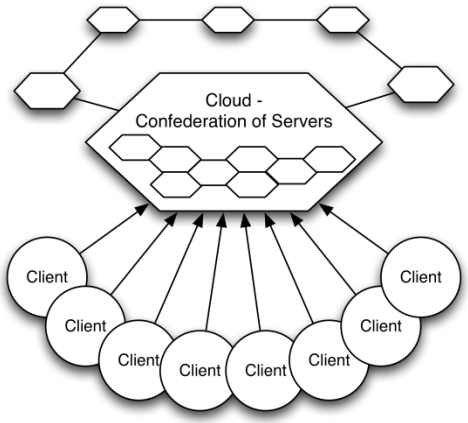
- Access to a distributed pool of resources
  - Resources: storage, bandwidth, computational power

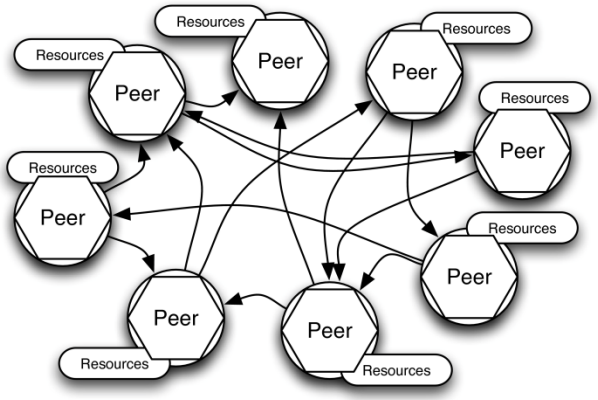
Cloud computing

- Resource providers: companies
- Controlled environment
  - No malicious providers
  - No (/minimal) churn
  - Homogenous devices
- Selective centralized structures
  - Accounting, monitoring
  - Single access point
  - Centralized updates

P2P systems

- Resource providers: user devices
- Uncontrolled environment
  - Churn, malicious providers
  - Heterogeneous devices
  - Uncertainty / unpredictability
  - Distributed access points





Essential Aspects of Cloud Computing

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

On-demand self-service

- resources (e.g., server time, network storage) are automatically provided to a customer when required

Rapid elasticity

- underlying infrastructure is able to adapt to changing requirements
  - (e.g., number of concurrent users)
- → allows for dynamic up-/down-scaling

Measured Service

- metering of resource and service consumption to provide elastic pricing and billing models
- e.g., pay-per-use

Resource pooling

- resources are provided/assigned dynamically in a multi-tenant way

Broad network access

- capabilities are available worldwide over standard network mechanisms

P2P vs. Cloud Computing – Cloud Service Models

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Software as a Service

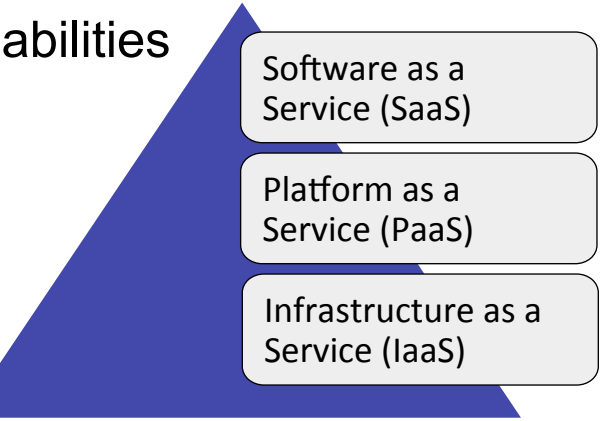
- provides applications / services representing business functions
- e.g., Google Docs, Salesforce CRM

Platform as a Service

- provides a platform for application / development and hosting
- e.g., Google App Engine,
  Windows Azure (Platform)

Infrastructure as a Service

- provides storage, computing and network capabilities
- e.g., Amazon S3, Amazon EC2, SQL Azure



Short Question: Why do we need P2P then?

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

P2P in the Cloud – Backend

- Decentralized mechanisms, data storage, big data ...

Costs

- Resources are available and for free

Security

- Integrate security in the mechanisms
- No central point to get data over all users
- Protocols must assume that everyone is an attacker

Locality / „Offline“ usage

- Local data exchange in companies / private households

Towards Distributed Systems

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Large set of resources available

- Resource in this lecture: content and capacity
- Types of content are
  - Programs, data, information …
- Capacities are
  - CPUs, storage, memory, bandwidth …

Observation

- Quality demand of users of the content grows
  - Processing speed
  - Amount / size of content
- Demand goes beyond the capabilities of single devices

“640 kB ought to be enough for anybody.”
- 1981, not Bill Gates

"I think there is a world market for about five computers”
- 1940s, IBM's president, Thomas J Watson

Trends: CPU power, Bandwidth

Moore‘s Law

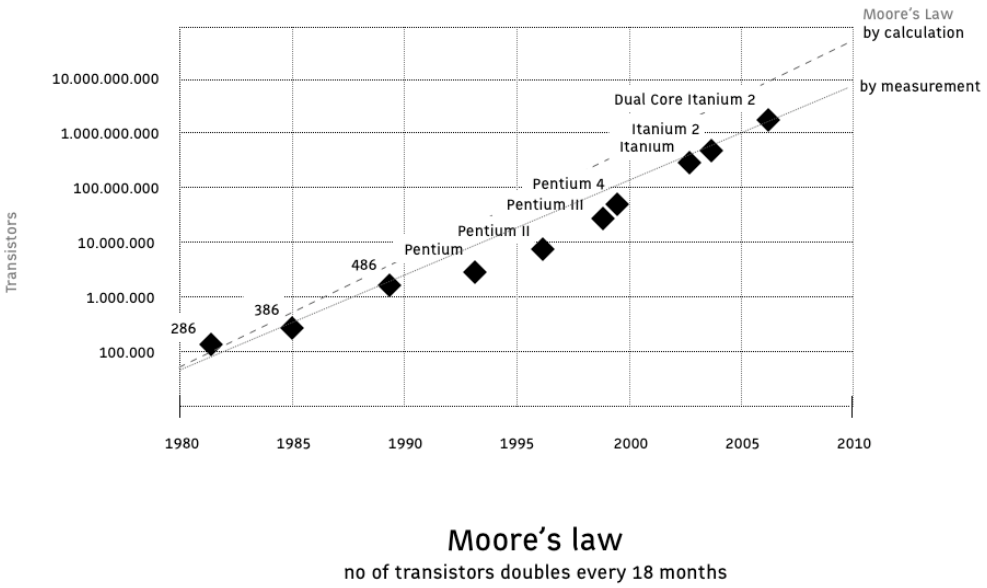
- Transistors on Chip doubles every 18 months

→ Increase of computational power

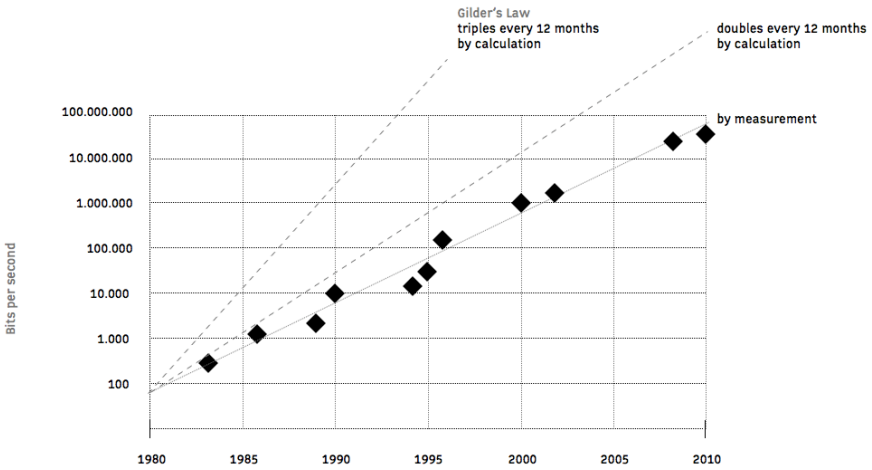
Glider‘s Law

- Bandwidth capacity triples every 12 months
- Also mobile Internet

→ Increase of Bandwidth



Moore's law
no of transistors doubles every 18 months



Gilder's law
bandwith capacity triples every 12 months
following a NIELSEN estimate

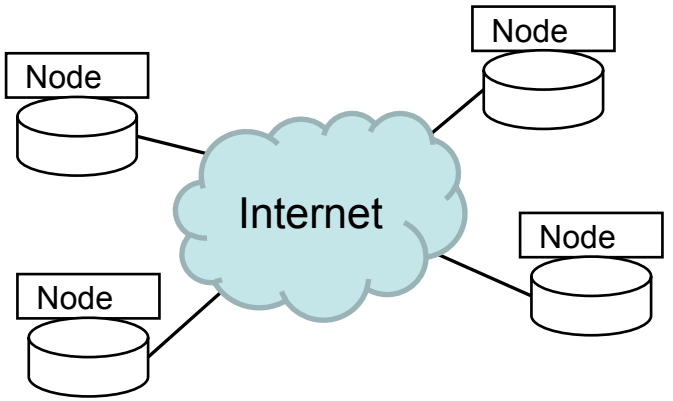
Motivation for Peer-to-Peer Networking

The Internet

- Routers and endpoints
- Endpoints: „Servers“ and private devices
  - Private devices are in majority

Trends

- Since ~ 2000
  - Private devices powerful enough to serve others
  - Compression leads to manageable content
    - mp3, mp4, divx, …
- User generated (+uploaded) content
  - Blogs, music, videos, software
  - Also: personal communication (voice, video)



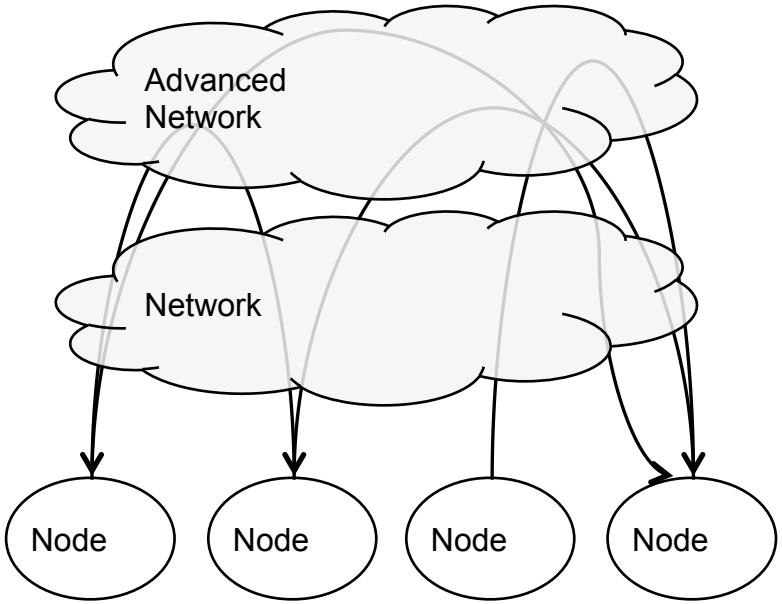
Main idea of Peer-to-Peer:

- Use (only) user devices to serve other users
- Self-organizing network of providers and consumers

Typical Service / Content Discovery and Provision

Intelligence in the network

- Enabling search for resources
- “Content”-based routing
- Provider and client matching
- All roles are distributed fulfilled by large number of nodes



Summary on the motivation

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

A huge number of nodes participating in the network

- Have resources to share
- Have demands towards the use of resources
  which may not be satisfied easily and by single nodes

??? Main question for “intelligent network”

- How to find nodes providing desired resources
- How to organize the exchange of resources

Peer-to-Peer (P2P)

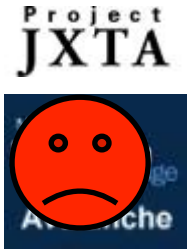
- P2P builds overlay network(s)
- P2P overlay offers mechanisms to find / look up what is wanted

Mode of operation

- After locating the node providing the desired service:
- Interact directly from peer to peer

Question: Which p2p applications do you know?

Success of P2P Networking



gnutella.com

Initially

- 40 million Napster users in 2 years → long living (BitTorrent)

Many applications dead or not p2p anymore

New applications coming

Above logos copied from the respective web page

P2P Traffic

P2P traffic was the major traffic source,

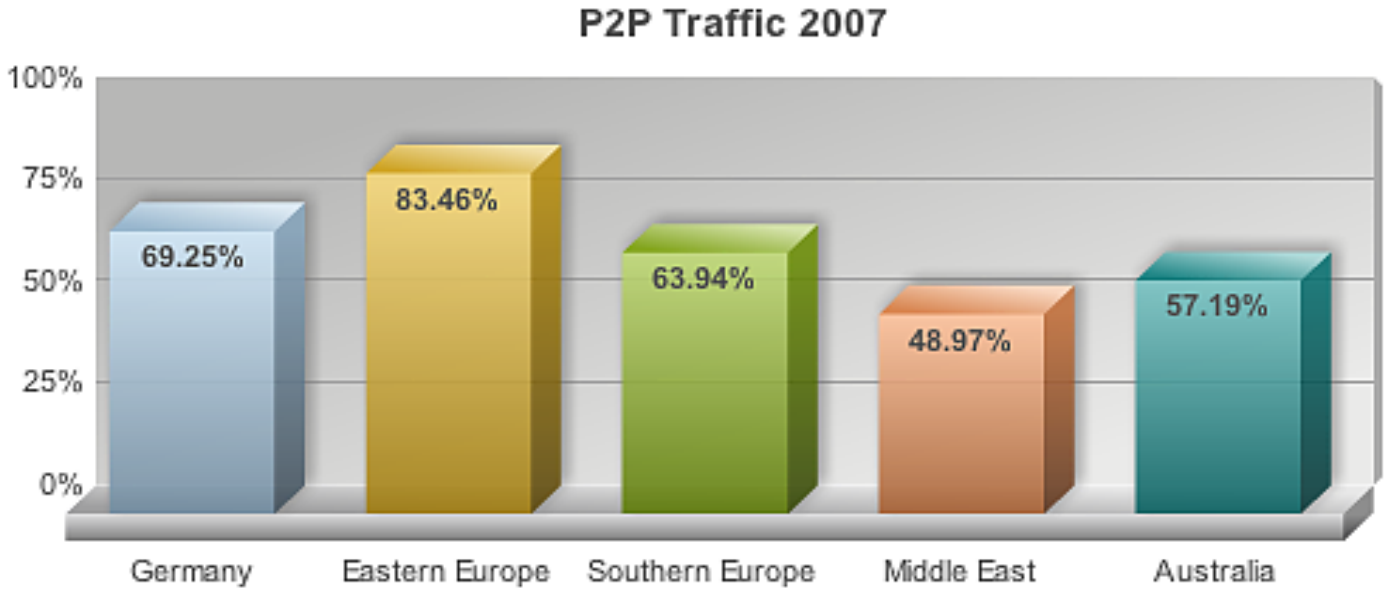
- since 2003: In Internet traffic more than ~50% is P2P traffic

P2P traffic in the Internet

- 60% – 80% P2P file sharing traffic on backbones

Recent measurements

- Video streaming (YouTube, Hulu, Netflics) dominates



Source: http://www.ipoque.com/resources/internet-studies

Studies on Internet Traffic 2013 – P2P Ratio

HEINRICH HEINE UNIVERSITÄT DÜSSELDORF

Table 10. Global Consumer Internet Traffic, 2012–2017

Consumer Internet Traffic, 2012–2017	2012	2013	2014	2015	2016	2017	CAGR 2012–2017
By Network (PB per Month)							
Fixed	25,529	32,097	39,206	47,035	56,243	66,842	21%
Mobile	684	1,239	2,223	3,774	6,026	9,131	68%
By Subsegment (PB per Month)							
Internet video	14,818	19,855	25,800	32,962	41,916	52,752	29%
Web, email, and data	5,173	6,336	7,781	9,542	11,828	14,494	23%
File sharing	6,201	7,119	7,816	8,266	8,478	8,667	7%
Online gaming	22	26	32	39	48	59	22%
By Geography (PB per Month)							
Asia Pacific	9,033	11,754	14,887	18,707	23,458	29,440	27%
North America	6,834	8,924	11,312	14,188	17,740	21,764	26%
Western Europe	5,086	5,880	6,804	7,810	9,197	10,953	17%
Central and Eastern Europe	2,194	2,757	3,433	4,182	5,015	5,897	22%
Latin America	2,656	3,382	4,049	4,588	5,045	5,487	16%
Middle East and Africa	410	640	944	1,334	1,816	2,432	43%
Total (PB per Month)							
Consumer Internet traffic	26,213	33,337	41,429	50,809	62,269	75,973	24%

Source: Cisco VNI, 2013

Studies on Internet Traffic 2013 – File Transfer

HEINRICH HEINE
UNIVERSITÄT DÜSSELDORF

Table 12. Global Consumer File-Sharing Traffic, 2012–2017

Consumer File Sharing, 2011–2016	2012	2013	2014	2015	2016	2017	CAGR 2012–2017
By Network (PB per Month)							
Fixed	6,155	7,060	7,738	8,159	8,343	8,506	7%
Mobile	45	59	78	107	135	161	29%
By Subsegment (PB per Month)							
P2P file transfer	5,374,262	5,981,677	6,330,010	6,404,161	6,199,877	5,893,411	2%
Other file transfer	826,343	1,137,158	1,486,266	1,861,915	2,277,919	2,773,901	27%
By Geography (PB per Month)							
Asia Pacific	2,358	2,801	3,156	3,440	3,655	3,857	10%
Central and Eastern Europe	911	1,079	1,242	1,374	1,427	1,494	10%
North America	829	982	1,090	1,173	1,244	1,302	9%
Western Europe	1,350	1,367	1,360	1,305	1,255	1,235	-2%
Latin America	666	782	850	862	807	725	2%
Middle East and Africa	88	108	118	112	89	53	-10%
Total (PB per Month)							
Consumer file sharing	6,201	7,119	7,816	8,266	8,478	8,667	7%

Source: Cisco VNI, 2013

Dominant P2P Applications

2003: Sandvine Study

- in Europe (France, Germany, ..)
  - predominant EDonkey/EMule
- in USA
  - predominant KaZaA/Fastrack

2005:

- BitTorrent
  - most successful file sharing P2P application
- Skype dominates in IP telephony
- KaZaA more and more irrelevant
- eDonkey replaced by eMule
  - using an extended but compatible protocol

2009:

- Wuala
  - P2P-based storage service
- KaZaA and eMule almost dead
- PPLive
  - P2P-based Video Streaming Platform
  - Mostly used in Asia
- BBC IPlayer
  - http://www.bbc.co.uk/iplayer/
- Vuze
  - Former Azureus
  - P2P-based Video-on-demand platform

>2014:

- Distributed social networks?
- P2P cloud?
- Local mobile data sharing