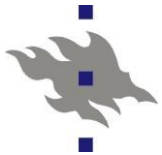
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Peer-to-Peer Networks

Chapter 1: Introduction

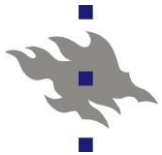
Jussi Kangasharju



Chapter Outline

- Course outline and practical matters

- Peer-to-peer (P2P) overview
- Definition of P2P
  - What is P2P and how it is different from other similar systems
- History of P2P
- Current state of P2P
- Why P2P works?
- Future of P2P
- Outline of future chapters

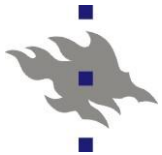


Course Details

- Lectures: Mondays 10-12 in B222 (Exactum)
  - Through Period III and IV
- Exercise sessions: Fridays 10-12 in C222 (Exactum)

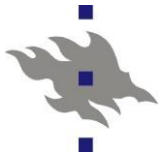
- Exercises both written and programming
  - Exercises can be done in groups
  - Exercises give bonus for exams (details in exercises)

- Two exams:
  - Thursday 28.2. 16-19 and Wednesday 30.4. 9-12
  - Both exams count for 50% of the grade



Course Practical Arrangements

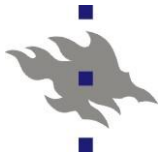
- Lecturer: Prof. Kangasharju
  - Office hours: Mon 12-13 and Fri 9-10 in D233
  - Other appointments by email
- Possible literature: Peer-to-Peer Applications and Systems
  - From Springer, Editors R. Steinmetz, K. Wehrle
  - General book about many P2P issues, some are covered in this course, others not
  - Can be downloaded for free from university network
- Course announcements on course web page
  - News, announcements
  - Slides
  - Exercises



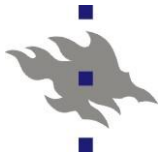
Course Outline and Goals

- Course topic is peer-to-peer networks and systems
- Take a look at current state in P2P systems, both in “real world” and in research work

- What does P2P mean?
- How is P2P principle reflected in actual systems?
- Why does P2P work?
- How to evaluate P2P systems?

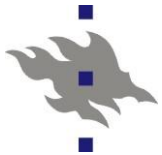


Questions?



Peer-to-Peer?

- What is it?
- What does the word “peer” mean?
  - Merriam-Webster: one that is of equal standing with another **:** EQUAL; *especially* **:** one belonging to the same societal group especially based on age, grade, or status
- Peer-to-peer: From one equal partner to another?
- How do you define peer-to-peer?

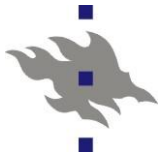


P2P Principle

- P2P can be seen as an organizational principle
  - System exhibits the P2P principle more or less clearly
- P2P principle applicable to many kinds of systems
  - Content distribution, communication, distributed computation, and collaboration

- Core concepts of the P2P principle:
  - Self-organizing, no central management
  - Resource sharing, e.g., files
  - Based on voluntary collaboration, e.g., Wikipedia
  - Peers in P2P are all equal (more or less)
  - Large number of peers in the network

- In contrast: Client-server systems have clearly defined roles for client and server

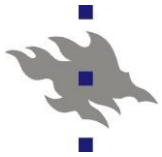


Definition of P2P

A P2P system exhibits the following characteristics:

1. High degree of autonomy from central servers
2. Exploits resources at the edge of the network
   - Storage, CPU cycles, human presence
3. Individual nodes have intermittent connectivity

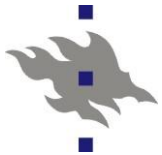
- Not strict requirements, instead typical characteristics
- Above characteristics allow us to distinguish P2P systems from other similar systems



Properties of P2P Systems

- P2P systems typically have the following properties:

1. Unreliable, uncoordinated, unmanaged
   - No central authority, peers are completely independent
   - Increases flexibility of individual peers, but makes the overall system (possibly) unreliable
2. Resilient to attacks, heterogeneous
   - Large number of peers in the system, hard to bring it down?
   - Heterogeneous peers make viruses and worms harder to write?
3. Large collection of resources
   - Voluntary participation, global reach
   - Millions of simultaneous users



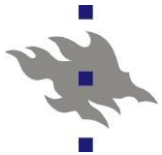
P2P Vision

- P2P vision for the future:

No More Dedicated Servers,
Everything in Internet Served by Peers

- No mail servers, no file servers, no web servers
- Individual peers, operating independently from one another offer all the basic services

- Is this a realistic vision?

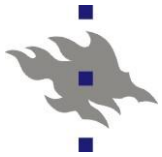


What is P2P?

- So, what exactly is P2P then?!?

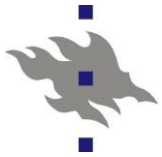
- Many peers distributed in the network
- So, P2P is same as distributed systems?
- Not quite…

- So, what are the relationships between P2P and
  - Distributed systems?
  - Grid computing?
  - Ad hoc networks?
  - (from the point of view of someone working on P2P… :-)



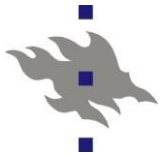
Distributed Systems

- Distributed systems (DS) around for long time
- Definition (FOLDOC)
  - *A collection of (probably heterogeneous) automata whose distribution is transparent to the user so that the system appears as one local machine. This is in contrast to a network, where the user is aware that there are several machines, and their location, storage replication, load balancing and functionality is not transparent. Distributed systems usually use some kind of client-server organisation.*
  - Note: In its loosest sense, distributed system is any system with several nodes and a network between them
- Above definition implies a managed and controlled network which acts as a single, logical computer
- P2P is not controlled, not managed, existence of several peers is not hidden



Grid Computing

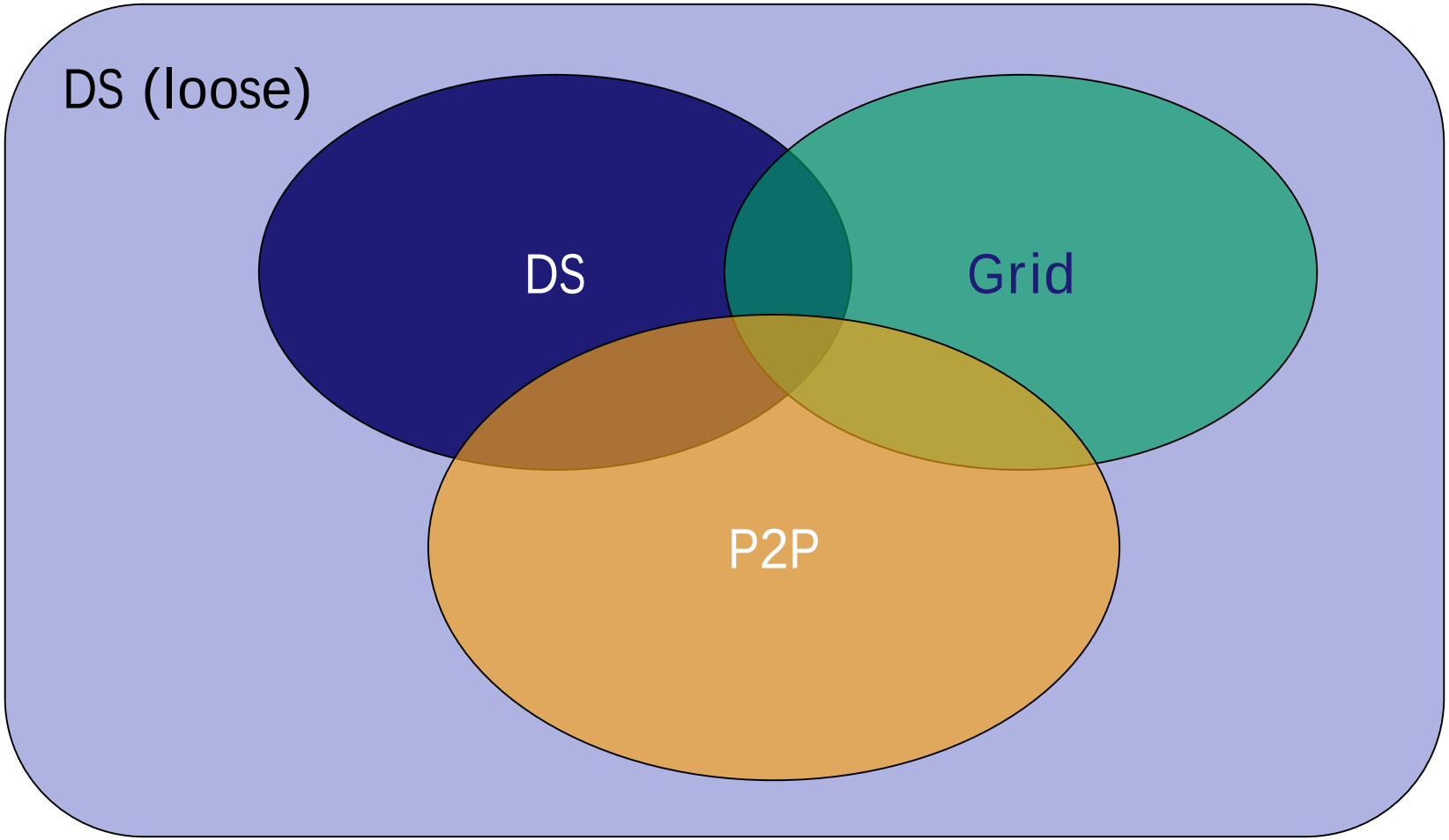
- Grid computing is a recent phenomenon
  - About 10 years since first coherent vision
- Grid definition: (I. Foster)
  - *Coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations.*
- Hence, grid is coordinated, aimed at solving problems
  - Motivation to join a grid: more resources available
  - Resources typically provided by everyone and usable by all
- P2P a bit similar, working towards a common goal
  - But P2P not coordinated, resources only provided by some?
  - P2P not based on institutions



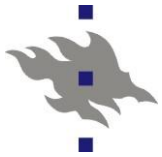
Ad Hoc Networks

- Wireless ad hoc networks becoming hot
  - Origins in military uses, later also emergency management
  - Commercial uses so far non-existent
- Typical characteristics:
  - No infrastructure available
  - Individual nodes provide resources for the network
  - Common goal
- Resources typically network bandwidth (routing)
- Goal of ad hoc networks: Enable communications
- In some ways, ad hoc networks can be seen as P2P
  - In this course, we do not go in details of ad hoc networks
  - Core issues in ad hoc networks quite different from P2P, main focus is on routing

P2P vs. DS vs. Grid

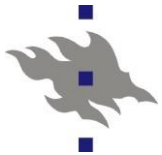


Note: Sizes of bubbles not to scale… ;-)



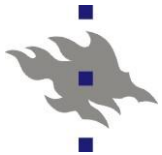
History of P2P

- What was the first P2P system and when?
- Answer: ARPANet 1969
- Later: USENET, 1979 (also FidoNet 1984, other BBSs)
  - Current Internet routing (BGP) is P2P
- The term P2P was coined by Napster in 1999
- Napster was a huge hit, brought P2P to general attention
- Illegal sharing of copyrighted material by users was the main driver behind Napster's success and the reason for its downfall
- Other systems followed Napster quickly, based on other design choices
- Research community followed suit quickly
  - Many deployed systems proprietary, hard to examine well…



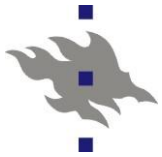
Current State of P2P

- Where are we now?
- P2P networks going strong, all over the world
  - Many networks highly popular and widely used
  - Different networks in different countries
- P2P networks currently mostly used for illegal sharing of copyrighted material
  - Music, videos, software, …
  - Note: Can be used for legal sharing too (see BitTorrent)
- Other applications starting to emerge (see below)
- Content providers not so happy
  - Sue companies making P2P software (e.g., Napster), sue software developers (Winny), sue users sharing material
  - But also providing alternate means: iTunes & friends



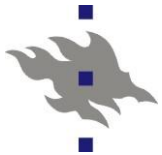
New P2P Systems

- File sharing was first P2P application
- Other applications are coming to light
- BitTorrent more content distribution than file sharing
- P2P extending beyond file sharing: Skype
  - We will look Skype closer in Chapter 2
- Skype is a P2P telephone “system”
  - Can call other computers, or normal phones
- Skype is based on the KaZaA network (see Chapter 2)
- Similar to VoIP services (e.g., Vonage), but fully based on the individual peers
  - Skype requires a computer, VoIP services often do not



P2P: Some Statistics

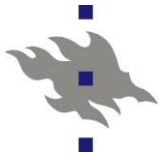
- Currently P2P accounts for 70% of network traffic
  - A bit different in different networks
  - Hard to measure accurately
- Network providers (ISP) not too happy about this
  - But: Often traffic internal to ISP! (e.g., T-Com)
- Some numbers: (take with a grain of salt…)
  - KaZaA had 60 million users total, 1-5 million online at any time
  - 85 million downloads/day
    - Software downloaded over 230 million times
  - Google has 80 million users/month, 200 million queries/day
  - Skype has over 200 million users, over 10 million concurrently



Current State in Research

- Lot of interest in P2P in the research world
  - Common to networking and distributed systems
- Strong focus so far on searching and locating objects in P2P networks
- Some work on replication, robustness, and security
- Higher level work on filesystems, P2P applications
  - See later chapters for examples

- Alas, P2P has become buzzword?
  - Owes to confusion about terminology and merging of different research communities
  - No commonly accepted definition of P2P

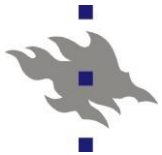


Why Does P2P Work?

- Why P2P file sharing networks are so successful?

1. Easy to use
   - P2P software readily available, simple to use
2. Provide something useful (for free)
   - Until recently, only alternative to P2P content was “buy a CD”
   - Online music stores may change this?
3. Anyone can contribute
   - Contributions not tied to geographical location; user in Brazil can provide files for everyone (compare with ad hoc networks!)
   - Enough “altruistic” users to make P2P networks useful

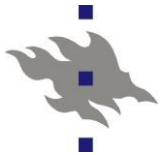
- Some systems (Skype) completely hide the P2P-part
  - Will this become the future trend?



P2P: Traps and Pitfalls

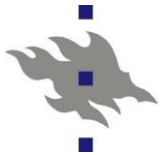
- What could render current P2P networks useless?
  - In particular, file sharing networks

1. Removal of desirable content
   - Stricter enforcement of copyright laws?
2. Alternative ways of getting same content
   - Online music stores?
3. Blocking of P2P traffic by ISPs
   - Or making users pay for bandwidth they use?
4. Viruses or worms on P2P networks
   - Exploit bugs in P2P software



When P2P and When Not P2P?

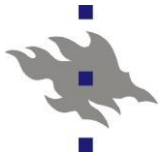
- So, when is P2P the right solution?
- Or, when is P2P the wrong solution?
- *Claim:* Our earlier P2P vision is technically feasible
  - In other words, possible to build everything on Internet without any dedicated servers
- Gotcha: Just because it's technically feasible, doesn't make it sensible…
- In other words, just because we can do it P2P, doesn't mean that we should do it P2P
  - True in many areas of life…
- So, when *is* P2P the right solution?!?



Some Criteria

- Let's consider the following criteria

1. Budget
   - How much money do we have?
2. Resource relevance
   - How widely are resources interesting to users?
3. Trust
   - How much trust there is between users?
4. Rate of system change
   - How fast does "something" in the system change
5. Criticality
   - How critical is the service to the users



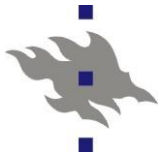
Analysis

Budget

- If you have enough money, build a centralized system
- Look at Google if you doubt this claim ;-)
  - Any system can be made to scale with enough money
- P2P is therefore useful when budget is not unlimited
  - In other words, most real-world situations…
  - From the rest of this analysis, we assume limited budget

Resource relevance

- If shared resources are highly relevant to a large number of users, P2P makes sense
- Easier to build a distributed solution when interest is widely spread



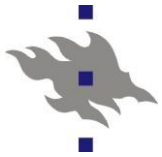
Analysis, Continued

Trust

- If other users can be trusted, P2P is a good solution
  - For example, corporate network or any closed network
- Building a fully distributed, trusted network is still very much a research problem (and may remain so…)

Rate of system change

- How high are the system dynamics?
  - Rate of peers joining and leaving, rate of information change in system, rate of change in network topology, …
- If the rate of change is too high, a distributed P2P solution might not be able to keep up
- Again, research problem



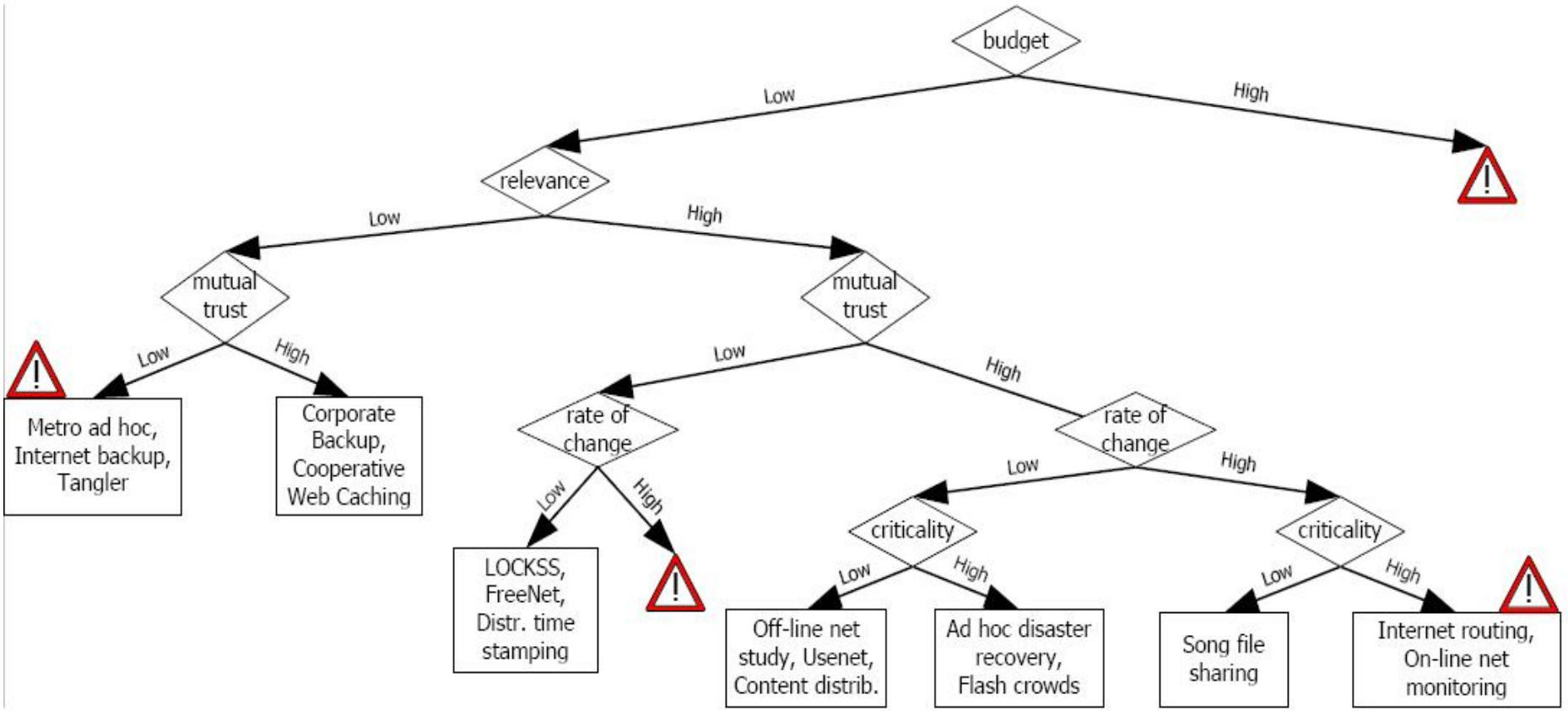
Analysis, End

Criticality

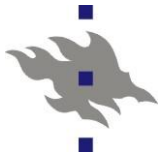
- How important is the service to the users?
- If you “can live without it”, P2P is acceptable
- If “it must work”, then consider other solutions…

- Summary: P2P is good when:
  - Budget is limited
  - Resources have wide interest and relevance
  - Trust between participants is high
  - Rate of change is manageable
  - Criticality is low
- Note: Again, no need to fulfill every point!

P2P Suitability Tree and Examples

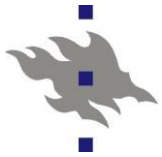


Taken from M. Roussopoulos et al. “2 P2P or not 2 P2P?”, IPTPS 2004



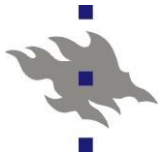
What does Future Hold for P2P?

- Take out crystal ball and look 5 years into future?
  - P2P has been around for a bit over 5 years now…
- Where will file sharing be in 5 years?
  - Still popular? Underground activity?
- P2P content distribution? (BitTorrent and others)
  - Microsoft building their system for software patches?
  - Some other systems patch via BitTorrent
- How about Skype and others?
  - Will Skype be around in 5 years?
  - Will Internet telephony be taken over by telcos?
- Research efforts in P2P?
  - More mature, concentrate on fundamental principles
  - What makes P2P different from other systems?



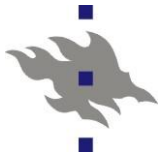
Future of P2P?

- Global P2P networks?
  - Besides file sharing, “Skype”, and research prototypes?
- Ubiquitous computing
  - Small, autonomous devices collaborating à P2P principle in action?
- For insight on future trends: Look what is going on in Korea now :-)
  - High bandwidth residential and wireless access
  - Online gaming (50% of network traffic!) main source of traffic
  - File sharing moved to pay models
  - Online communities gaining importance
  - Of course, cultural differences can change things…



Chapter Summary

- Peer-to-peer principle of self-organization and resource sharing
- P2P systems exhibit following characteristics:
  - Autonomy from central servers
  - Use of edge resources
  - Intermittent connectivity
- Hard to define clearly the limits of P2P
  - Compare with distributed systems and grid computing
  - Different people working in different areas have different definitions



Outline of Future Chapters

- Chapter 2: Current P2P Systems
- Chapter 3: Networks, Searching, and DHTs
- Chapter 4: P2P storage
- Chapter 5: Reliability and performance in P2P Networks
- Chapter 6: P2P Content Distribution
- Chapter 7: P2P Other Issues