Overlay and P2P Networks

Introduction

PhD Samu Varjonen
18.1.2016
Contents

• Course Overview
• Lectures
• Assignments/Exercises
Course Overview

• Overlay networks and peer-to-peer technologies have become key components for building large scale distributed systems.

• This course will introduce overlay networks and peer-to-peer systems, discuss their general properties, and applications. The course will cover the following topics:
  – Currently deployed peer-to-peer systems and how they work
  – Distributed Hash Tables as a base for structured peer-to-peer systems
  – Peer-to-peer storage systems
  – Performance issues, legal aspects, and privacy issues
  – Overlay content distribution algorithms
General Info

Advanced course, 5 credits

Requirements: basics of networking

Assignments/exercises done as group work (1-2 persons), idea is to keep the same group structure (but do tell about freeriders!)
Lectures

• Lectures
  – Monday 14-16 in D122 18.1.-3.3.
  – Thursday 12-14 in D122 18.1-3.3.

• Assignments/exercises
  – Wednesday 12-14 B222 18.1-4.3.

• Course based on book
Exam material: additional articles


Lectures

18.1. Introduction. Exercises.
20.1. Unstructured networks I
25.1. Unstructured networks II
28.1. Bittorrent, modelling and evaluation
1.2. Freenet and intro to power-law networks.
4.2. Consistent hashing. Distributed Hash Tables (DHTs)
8.2. DHTs continued.
15.2. Power-law networks and applications.
18.2. Applications I
22.2. Applications II
25.2. Advanced topics
29.2. Conclusions and summary
Grading

Course grading will be based on the final exam and the assignments/exercises.

Course exam 9.3.2016 16:00 B123

Separate exam on 22.4.2015 16:00 in B123
    You can use the exercise score with the first separate exam
Assignments/Exercises

- Assignments are done in two-person groups (or alone), groups can change between assignments
- Assignments give bonus points for the exam
  - Max exam points 18, max bonus points 4
- Wednesday 12-14 B222 last session 2.3.
  - 20.1. Reception on questions I
  - 26.1. Answers due to questions I
  - 27.1. Feedback on questions I
  - 3.2. Reception on questions II
  - 9.2. Answers due to questions II
  - 10.2. Feedback on questions II
  - 17.2. Reception on questions III
  - 23.2. Answers due to questions III
  - 24.2 Feedback on questions III
  - 2.3. Generic questions and answers, to prepare for exam.
Consistent hashing alleviates network problems and eventual consistency can be achieved through replication and synchronization

Examples: Dynamo, Cassandra

Replication, Gossip, etc.

Selective flooding

Consistent hash (O(1) DHT)

Good for arbitrary data and search functions, can aggregate routing info, structure improves scalability

Examples: Gnutella and Freenet

Example: BitTorrent

Limited flooding / depth first / Bloom filters

Tracker

DHT

Example:
- Lookup: Chord, CAN, Kademlia
- Storage: PAST
- Rendezvous: Scribe (for multicast), i3

Good for name/value data, note flat address space, one node is responsible, churn is a concern

Cluster

Wide-area (unstructured)

Search

Storage

Rendezvous

Search

Storage

Rendezvous

Search

Storage

Rendezvous

Wide-area (structured)

Internet (TCP/IP)
<table>
<thead>
<tr>
<th>Main theme</th>
<th>Prerequisites</th>
<th>Approaches learning goals</th>
<th>Meets learning goals</th>
<th>Deepens learning goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlay and peer-to-peer networks: definitions and systems</td>
<td>Basics of data communications and distributed systems (Introduction to Data Communications, Distributed Systems)</td>
<td>Knowledge of how to define the concepts of overlay and peer-to-peer networks, and state their central features</td>
<td>Ability of being able to compare different overlay and p2p networks in a qualitative manner</td>
<td>Ability to give one’s own definition of the central concepts and discuss the key design and deployment issues</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ability to describe at least one system in detail</td>
<td>Ability to assess the suitability of different systems to different use cases</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ability of being able to compare different overlay and p2p networks in a qualitative manner</td>
<td>Ability to give one’s own definition of the central concepts and discuss the key design and deployment issues</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ability of designing distributed hash table-based applications</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Knowledge of key performance issues of distributed hash table systems and the ability to analyze these systems</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distributed hash tables</td>
<td>Basics of data communications and distributed systems (Introduction to Data Communications, Distributed Systems)</td>
<td>Knowledge of the concepts of structured and unstructured networks and the ability to classify solutions into these two categories</td>
<td>Ability of being able to compare different distributed hash table algorithms</td>
<td>The knowledge of choosing a suitable distributed hash table design for a problem Familiarity with the state of the art</td>
</tr>
<tr>
<td></td>
<td>Basic of Big-O-notation and basics of algorithmic complexity</td>
<td>Knowledge of the basics of distributed hash tables</td>
<td>Ability of designing distributed hash table-based applications</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ability to describe at least one distributed hash table algorithm in detail</td>
<td>Knowledge of key performance issues of distributed hash table systems and the ability to analyze these systems</td>
<td></td>
</tr>
<tr>
<td>Reliability and performance modelling</td>
<td>Basics of probability theory Basics of reliability in distributed systems</td>
<td>Ability to model and assess the reliability of overlay and peer-to-peer networks by using probability theory</td>
<td>Ability of analytically analyzing the reliability and performance of overlay and peer-to-peer networks</td>
<td>Familiarity with the state of the art</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Knowledge of the most important factors pertaining to reliability</td>
<td>Understanding of the design issues that are pertinent for reliable systems</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ability to assess the reliability of overlay and peer-to-peer networks by using probability theory</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content distribution</td>
<td>Introduction to Data Communications</td>
<td>Knowledge of the basic content distribution solutions</td>
<td>Knowledge of different content distribution systes and the ability to compare them in detail</td>
<td>Familiarity with the state of the art</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ability to describe at least one overlay and p2p network based content distribution solution</td>
<td>Knowledge of several content distribution techniques</td>
<td></td>
</tr>
<tr>
<td>Security</td>
<td>Basics of computer security</td>
<td>Knowledge of the basic security issues with overlay and p2p networks Knowledge of the sybil attack concept</td>
<td>Ability to discuss how security problems and limitations can be solved Knowledge of how to prevent sybil attacks</td>
<td>Knowledge of how to prevent sybil attacks Familiarity with the state of the art</td>
</tr>
</tbody>
</table>
Contact information

Lecturer PhD Samu Varjonen (contact info on homepage)

Assignments: Juhani Toivonen (@cs.helsinki.fi)

Course homepage can be found: www.cs.helsinki.fi/courses
Questions?
Overlay and P2P Networks

Introduction and unstructured networks

PhD Samu Varjonen

18.1.2016