Exercise
for Lecture "P2P Systems"

Prof. Dr. David Hausheer

Dipl.-Wirtsch.-Inform. Matthias Wichtlhuber, Leonhard Nobach, M. Sc., Dipl.-Ing. Fabian Kaup, Christian Koch, M. Sc., Dipl.-Wirtsch.-Inform. Jeremias Blendin

Surname (Nachname):	
First name (Vorname):	
ID# (Matrikelnummer):	

Problem 1.1 - RB-HORST User Study

RB-HORST is a research prototype resulting from a research project sponsored by the European Union. We will utilize a software prototype as a part of the P2P lecture's exercise to allow students to have a hands-on experience with a live deployment of a cutting-edge distributed Peer-to-Peer system and the respective hardware in their home premises.

- **Please take part in the following survey:** http://bit.ly/rbhsurvey

Problem 1.2 - P2P Architectures and Classifications

A) Argue why it is harder for an authority to shut down a decentralized P2P system than to shut down a Client-Server system. How about centralized P2P networks?
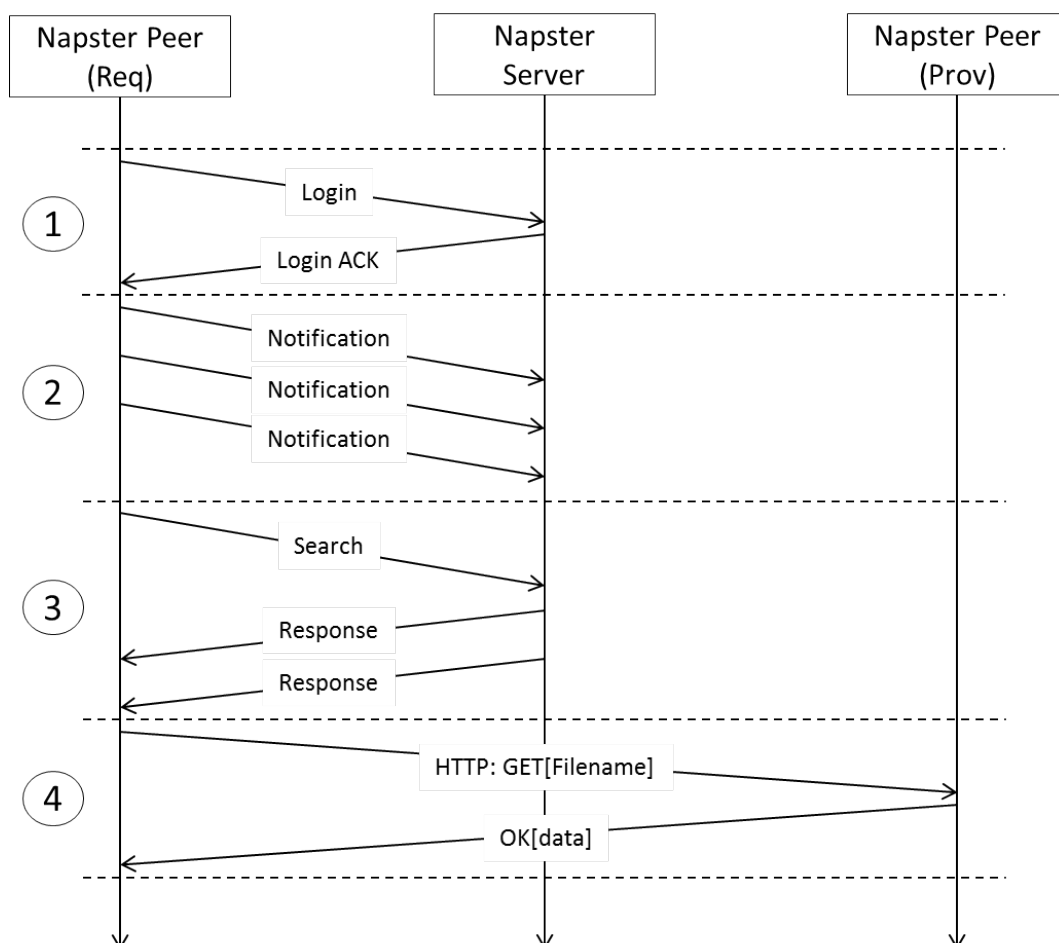
B) Choose the right answer:

		TRUE	FALSE
i)	In a DHT-based P2P network the connections in the overlay are "fixed".	☐	☐
ii)	A P2P system is more fault-tolerant than a Client/Server system.	☐	☐
iii)	A Client/Server system scales better with the number of users than a P2P system.	☐	☐
iv)	In a hybrid P2P network any terminal entity can be removed without loss of functionality.	☐	☐
v)	A hybrid P2P network suffers from a single point of failure.	☐	☐

Problem 1.3 - Napster

A) Does Napster match the derived key characteristics of P2P systems as defined in the lecture? Explain your answer.

B) Describe what is happening during phases (1) - (4) in the following sequence diagram of a Napster session. What problem can occur in step (4)?

Napster Peer (Req)	Napster Server	Napster Peer (Prov)

(1) Login → Server
Login ACK ← Server

(2) Notification → Server
Notification → Server
Notification → Server

(3) Search → Server
Response ← Server
Response ← Server

(4) HTTP: GET[Filename] → Prov
OK[data] ← Prov

Problem 1.4 - Gnutella

A) Which information (about the system) is required to configure the **TTL** field in Gnutella-like protocols? Consider the tradeoff of the probability to find all the potential matches in the network and the incurred overhead.

B) Which mechanism is used (besides TTL and Hop counter fields) by Gnutella to avoid loops while forwarding messages?

C) Given an origin node A in a Gnutella system, derive a formula $f(n, t)$ for the maximum number of reachable users from this node, given n, the number of neighbors per node, and t, the used TTL counter. Assume that no duplicate nodes are traversed on the path.

Use the formula to calculate the number of reachable users for t = 8 and n = 5 as well as for t = 7 and n = 8.

D) Derive a formula $g(n, x, y)$ for the maximum number of reachable users that are at least x but no more than y, with $x \leq y \leq t$, hops away from node A, using the assumptions of the previous task.

Calculate the number of reachable users that are between 6 and 8 hops away from A, assuming $n = 6$.

Problem 1.5 - Distributed Hash Tables

A) In a DHT, why is it important that node and data IDs are (nearly) random and equally distributed?

B) Name two advantages of unstructured (flooding-based) P2P architectures over structured ones (such as DHTs)?

C) Why do maintenance operations in DHTs (like Chord or Pastry) have a complexity of $O(log^2(n))$ but lookup operations only $O(log(n))$?

D) Explain why fuzzy queries are not simple to be implemented using Distributed Hash Tables?