Prof. Dr. Thorsten Strufe
Stefanie Roos
{strufe,roos}@cs.tu-darmstadt.de

Fachgebiet Peer-to-Peer Netzwerke
Fachbereich Informatik
Technische Universität Darmstadt

Exercise 3

for **Peer-to-Peer Networks** - winter term 2012/2013

(06.11.2012)

Deadline for submission: 13.11.2012 1:30 PM

Guidelines

- Exercises annoted by **G#** are intended to be discussed and solved in class without grading, whereas exercises annoted with **H#** are supposed to be solved in groups and be handed in for grading.
  This does not mean, that ungraded exercises are less important.
- Please submit your solutions until the beginning (1:30 PM) of the next exercise in the coming week or beforehand in S2|02-A312 on Monday, 11-12 AM or Tuesday, 9-12 AM. Electronic submission is also possible, until our post box is repaired. Please only use PostScript or PDF. We must be able to print out your submissions on a standard monochrome printer so that they are fully readable.
- Note that points are only given if your solution is clearly legible. Unreadable submissions will not be rated! (Machine written submissions are allowed)
- Written assignments are to be solved in groups of 2-3 participants while programming assignments have to be done in groups of four to six participants.
- Always annotate your solutions on the handed in sheet with names and matriculation numbers. If you have privacy concerns, you are allowed to omit your name and tell it to us personally.
- Please subscribe to the mailing list:
  `https://mail.rbg.informatik.tu-darmstadt.de/mailman/listinfo.cgi/p2p-lecture-ws12`
- By submitting any processed exercises or program code you hereby commit to the "Grundregeln der wissenschaftlichen Ethik am Fachbereich Informatik" (see also `http://www.informatik.tu-darmstadt.de/de/sonstiges/plagiarismus/`).
  This especially means, that you should always write in your own words. However if you use external materials, you have to cite them correctly.
  We will not accept solutions that only rely on literal citations.

G#3.1 Gnutella

To answer the following questions, assume the topology of a Gnutella network shown in Figure 1 where all peers maintain state information about messages they have already sent or received. Suppose that all messages are sent and received in a synchronized, round-based manner.

a) Query Routing in Gnutella

Peer 1 attempts to search for "snafu" using a TTL of 3. Note down all QUERY messages that are sent throughout the network until the TTL expires. Use the message format as shown by Table 1. Also note down the respective round and the TTL attached to each message.

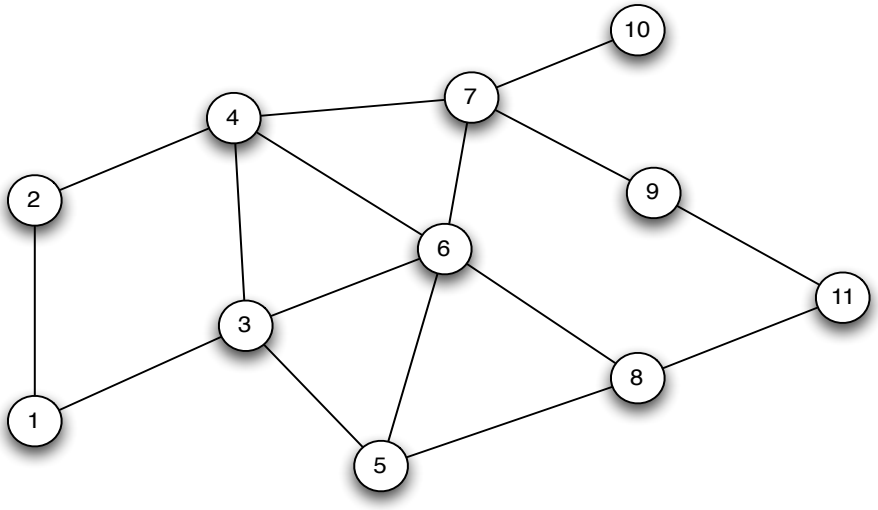

Figure 1: A Typical Gnutella Network

Table 1: Message format

Round	Sender	Receiver	TTL
...	...	...	...
...	...	...	...
...	...	...	...

b) Queryhit Routing in Gnutella

Assume that peers 2, 5, and 11 have files matching the keyword “snafu”. List all QUERYHIT message that are sent after the search, described above, has been started. Use the format shown in Table 1 without the TTL column.

c) State information

What would happen during the query routing if the peers would not store information about messages that they already received? Describe the problem using an example in the topology given in Figure 1.

d) TTL-Counter

Which TTL has to be chosen to be sure that every hit for a query is found in the topology of Figure 1? Which graph metric does this correspond to?

Is it sensible to choose the TTL so that all hits in a network are found in general?

G#3.2 BitTorrent prerequisites

If you are not sitting in class right now, please have a look on the slides that where presented in class and perhaps read the paper "Incentives Build Robustness in BitTorrent" by *Bram Cohen* from May 2003 to read up on to the BitTorrent protocol.

H#3.1 Bittorrent measurement and analysis (15 P.)

Read the Paper "The Bittorrent P2P file-sharing system: Measurements and Analysis" by *Pouwelse, Garbacki, Epema* and *Sips*, obtainable on the courses website.

Answer the following questions:

a) What are requirements of a good file sharing system?

b) Explain the different parties in Bittorrent and their connection.

c) How is a file published?

d) Define the terms availability, data integrity/pollution, flashcrowds and download performance as used in this paper. What do the measurements of this paper show with regard to Bittorrent?

e) What is a problem of big flashcrowds that is raised up by this paper? What mechanisms could be or are implemented to defeat that problem?

f) Give one reason why measurements are used to analyze P2P systems?

g) Discuss the advantages and disadvantages of a central data index and tracker as usually used for the BitTorrent protocol in comparison to ad-hoc querying inside the network as used in Gnutella.

h) What are newly introduced countermeasures to deal with the problem of a central tracker?