



1st Work Sheet Praktikum Protokolldesign WS 07/08

Question 1: (100 points) *Logfile analysis*

In the networking field one often has logfiles with sizes of several gigabyte. When analysing such logfiles one has to do it “intelligently”, since these files are too large to fit completely into main memory.

The file `/afs/net.t-labs.tu-berlin.de/home/praktikum/1.uebung/logfile.gz` contains the log file of a Web proxy. The logfile contains one line per access in the following format:

```
time duration remotehost code/status bytes method URL rfc931 peerstatus/peerhost type
```

- Write a Perl script¹ that calculates the data throughput for each access and outputs lines of the form “`size throughput`” to a file. Sort the file by increasing sizes.
- Write another script that outputs a similar file but converts the values of both columns to a logarithmic scale.
- Plot the results from (a) and (b) using `gnuplot`. Describe and explain the differences between the two plots. Try to find some reasons for the things you see in the plots.

You may use our template at `/afs/net.t-labs.tu-berlin.de/home/praktikum/1.uebung/gnuplot_template.plot` — just change the filenames and labels respectively, and call `gnuplot -persist your_modified_template.plot`.

Submission details: look at the FAQ

(on http://www.net.t-labs.tu-berlin.de/teaching/ws0708/PD_labcourse/)

Due Date: Tuesday, October 30, 2007, 11:59h s. t.

Notes:

- You can find further information online at http://www.net.t-labs.tu-berlin.de/teaching/ws0708/PD_labcourse/. Check the FAQ there. It contains details for submitting your solutions and questions concerning the work sheets. New information is added to the FAQ regularly.
- Please make sure, that mails to the address you gave upon registration are correctly delivered and regularly read by you.
- You can combine perl, shell scripts, unix tools (e.g., `sort`, `cut`, `uniq`, ...), etc. to solve the exercises. Quite often that’s the easiest solution. If you use such tools add a text file or script explaining how you combined the tools.

¹You may use C, C++, Java or **after consultation with us** another language instead — but believe us: you don’t want to do that :-)