

# Learning to use tools for measurement and analysis

Markus Peuhkuri

2020-09-15

First, see [introduction video](#) (available with Aalto credentials and review supporting [Linux Introduction](#) for examples, guidance and [code fragments](#).

After that, process following tasks.

This work has 3 tasks:

1. [Task 1: Process CSV data on command line](#)
2. [Task 2: Make a script to produce ping statistics](#)
3. [Task 3: Measure network throughput performance](#)

For each task, complete exercise and write report. In addition to task-specific questions, always describe your solution including samples of produced data (few lines). You can add scripts / programs as zip archive (submission is instructed separately).

Always make sure you have included **all details** in your answers and have answered **every** item.

You can perform tasks 2-3 also with your own computer (real or virtual) but note that the computer may not enter sleep state between instances. Of course, you can also download the file in task 1, too.

## Task 1: Process CSV data on command line

In this task you need to compute statistical values from a large (462 MiB) CSV file.

Aalto Linux computers have the course folder in `/work/courses/unix/T/ELEC/E7130/`. Under that folder is a directory `general/trace/tstat/2017_04_11_18_00.out/` that has a file `log_tcp_complete`. It is a space-separated CSV file with 130 columns and 886467 records. The first line is the header. Provide following answers (*in addition to description of your solution*):

1. How you can peek on file if it is too large to fit into memory?
2. Averages of columns 3, 7, 10, 17, 21, 24

3. Percentage of records where  $c_{10}/c_7$  exceeds a) 0.01, b) 0.10, c) 0.20 ( $c_{10}/c_7$  means for each line value in column 10 is divided by value in column 7)
4. Percentage of records where  $c_{24}/c_{21}$  exceeds a) 0.01, b) 0.10, c) 0.20
5. Largest values in columns 3, 9, 17, 23, 31

## Task 2: Make a script to produce ping statistics

Use script (and crontab or other timing method) to send 20 pings every 10 minutes to `ok1.iperf.comnet-student.eu` over period of 12 hours. You may want to synchronise this measurement with following one.

For the answer produce (*in addition to description of your solution and samples*) calculated for each hour:

1. Percentage of lost packets.
2. Average of successful round-trip-time (RTT) measurements.
3. Median of RTT. Lost packet is counted RTT as inf.

## Task 3: Measure network throughput performance

Use script (with crontab or alternative) to run iperf3 once a hour against `ok1.iperf.comnet-student.eu` over period of 12 hours, 10 seconds run each direction (minimum 24 runs total). You can run it from own computer or you can use Aalto servers.

As the iperf3 server supports only one session at time, the runs will be distributed according to following formula:

- minute: your student number modulo 60
- port: random selection of ports from 5200 to 5210 (inclusive, random for each run)

Produce an CSV file including a record for each run including columns:

- timestamp
- time of start (in hour:minute:second)
- total bytes transferred
- bitrate
- number of TCP retransmissions

Include sample of CSV file, tell how you produced the CSV data and graph.

For the answer produce (*in addition to description of your solution and sample of CSV file*):

1. Average bitrate for all measurement.
2. Minimum, median and maximum bitrate.

3. Graph comparing bitrate and number of TCP retransmissions.
4. If you used your own computer to run iperf, describe network connectivity.  
Can you make any conclusions of stability based on data?