Exercise 1 - Packets for S-38.3183 - Spring 2007

Mika Ilvesmäki
Helsinki University of Technology
Networking Laboratory
P.O. Box 3000
02015 TKK
Finland
mika.ilvesmaki@netlab.tkk.fi
s383183-exercise@netlab.tkk.fi

Abstract

Deadline for this exercise is 26.3.2007. To get your report into the grading process it must be submitted via email to *s383183-exercise@netlab.tkk.fi*. No returns in paper format or otherwise are allowed. The reports must be in pdf-format and sent as attachments to the email message. Do not forget your name and study book number from your report.

I. Introduction

This first exercise introduces you to packet trace files and how to get information out of them. The main learning goal of the exercise is to get you acquainted working with packet trace files and familiarize yourself with spatial and temporal aggregation.

II. Problems

All your answers should also contain the scripts/command line with which you created your answers. Lengthy entries of code should be put to the appendix of your exercise report. Remember always to comment (discuss) the results. The grading is largely based on your discussion of the results.

A. Traces

The traces are available from
http://www.netlab.tkk.fi/opetus/s383183/k07/exercises/traces/
For this exercise, choose one of the dec$x$-files, where x is

$$x = 1 + \text{Your study book number} \mod 4 \tag{1}$$

and choose one of the ebb$y$-files where y is

$$y = 1 + \text{Your study book number} \mod 3 \tag{2}$$

Remember to read the *readme.txt* -file. Note also, that the original files are big (and authentic) so you may need additional space in your work directory (use scratch- and tmp -directories). Catenate from the compressed files when possible.

III. Exercise questions

A. Basic packet trace data

For both of your traces determine the following basic properties:

- Length (in seconds) of the trace
- Number of packets in the trace
- Amount of data (in bytes) seen in the trace
- Number of flows in the trace
- Number and shares of TCP and UDP packets
- Number and shares of TCP and UDP data
- Number and shares of TCP and UDP flows

Present your results for the two traces in a table format.

B. Temporal aggregation - Time series

The intention of this part of the exercise is to introduce the concept of temporal aggregation. The idea is to aggregate data (packets and bytes) arrivals by observing these arrivals per certain time intervals (1 second, 10 second and 60 second intervals).

For both of your traces calculate the following timeseries:

- The number of packet arrivals per 1, 10 and 60 seconds. (Three timeseries in total.)
- The amount of data byte arrivals per 1, 10 and 60 seconds. (Three timeseries in total.)

Note: In order to complete this part of the exercise you are required to produce working code (in a programming language of your choice; Perl is recommended). You should include the code as appendix in your exercise report.

Present both timseries in figures and show in a separate table the mean value and the standard deviation of your timeseries. Try to make it relatively easy to compare the traces and timeseries. Furthermore, choose a method of moving average, present it (the method) and plot the moving average in the figure for all aggregation levels (1s, 10s, 60s). Minimize the number of figures you include in your report. However, do not sacrifice clarity!

Store both your timeseries data and the scripts you used to produce the data for the duration of this course.

C. Spatial aggregation - IP, protocol and port data

For both of your traces:

- List the Top 10 SrcIP-identifiers that show the most data (bytes) sent.
  How much of the total traffic does the Top 10 produce (show individual SrcIP-identifier contribution)?
  Tip: Aggregate by the SrcIP-identifier and for all identifiers add the bytes seen. You should end up with a list of SrcIP-identifiers and respective amount of data sent.
- Plot the cumulative distribution of data sent from each SrcIP-identifier. How much of the traffic (databytes) is sent by the top 10% of the most active senders? *Tip: Use cumsum-function in Matlab.*
- List the Top 10 TCP source ports and Top 10 TCP destination ports that show the most databytes.
  How much of the total traffic does the Top 10 produce (show individual TCP/Sport contribution)?
- The cumulative distribution of data sent from each TCP source port.
  How much of the traffic (databytes) is sent by the top 10% of the most active TCP source ports?

Use the code you produced in the previous subsection. Include the changes you (possibly) made to the code in the appendix of your exercise report.

D. Return of this exercises

Deadline of this exercise is March 26th, 2007. To get your report into the grading process it must be submitted via email to *s383183-exercise@netlab.tkk.fi*. No returns in paper format or otherwise are allowed. The reports must be in pdf-format and sent as attachments to the email message. Do not forget your name and study book number from your report.

IV. Acknowledgements

The author would like to thank CSC - the Center of Scientific Computing in Finland - for providing access to Funet network and for computing and archive resources and Lic. Sc. Markus Peuhkuri for his kind help in preprocessing some of the traces.