Exercise 3 - Correlations 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 10.4.2007. 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 third exercise introduces you to determining correlations within and between flow arrival processes. You are already familiar working with trace files and therefore the main learning goal of this exercise is getting acquainted with correlation analysis.

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$y$-files where y is

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

In addition you will have access to eight files of different types of email traffic:

- f0src25flw-timeseries.tab.gz
- f0dst25flw-timeseries.tab.gz
- f1src25flw-timeseries.tab.gz
- f1dst25flw-timeseries.tab.gz
- f0src993flw-timeseries.tab.gz
- f0dst993flw-timeseries.tab.gz
- f1src993flw-timeseries.tab.gz
- f1dst993flw-timeseries.tab.gz

These files contain timeseries that indicate the number of flows created at one second intervals.
The files are named as follows:

- *f0 and f1* indicate links. The measurements are made simultaneously on two parallel links. However, the direction of the traffic in the links is opposite.
- *src and dst* indicate whether a source or destination port has been chosen as aggregation point.
- *25 and 993* indicate the chosen port numbers
- *flw* indicates that activated flows are counted
- *timeseries* indicates that the file contains a timeseries

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. Autocorrelation in packet arrival process

As you may remember your dec$x$-trace contains information on packets. Determine bytes/s arrivals for the dec$x$- packet trace. You may choose the aggregation period for the bytes/s as you see fit. Values like 1s and 10s are recommended. Determine the following for the arrival process:

- Auto-correlation within one trace: Determine the auto-correlation values for bytes/s with increasing values of the lag. Plot the timeseries, and the autocorrelation values into the same plot. Discuss your findings.

B. Cross-correlation in application flows

You have access to a set of ready-made timeseries. Determine the cross-correlation (if any) between 25-smtp flows and 993-imaps flow arrivals.

- Pick two timeseries. Your choice of timeseries is important and effects the results.[1] Make an educated choice that is most likely to result in meaningful results. Justify, in your report, your choice of traces. You may compare several traces and report on all of your comparisons.
- Determine the cross-correlation (between the series) with increasing values of the lag. Plot both of the timeseries, and the cross-correlation values into the same plot. Discuss your findings.

[1] ...and grading!

C. Return of this exercises

Deadline of this exercise is April 10th, 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 the traces.