Exercise 3
for S-38.3183 - Spring 2006

Mika Ilvesmäki

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

Abstract. Deadline for this exercise is 9.4.2006. 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.

1 Introduction

The third exercise lets you determine correlations within and between the trace files.

2 Problems

All your answers should also contain the scripts/command line with which you created your answers. Remember to comment (discuss) the results. Please note that the data collection is event based, whereas many of the questions require you to solve a property for a fixed time interval (packets/s). Therefore, as a data preprocessing task, you have to determine the appropriate values before applying analysis methods.

Suggestion: Suppose you have to determine the packet/s for a trace in one-second intervals. One option would be to count the amount of packets between 0 and 1s; 1s and 2s etc. For a 60 minute trace you would get 3600 seconds with packet arrivals/count determined for each second. You could also experiment by determining the property with, say, 0.5s or 5 s intervals.

2.1 Traces

The traces are available from
http://www.netlab.tkk.fi/opetus/s383183/k06/exercises/traces/
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 zip-files when possible.

2.2 Correlation in packet arrival process

Determine packet/s arrivals for two *different* packet traces (dec2 and dec3) with *the same timeout* (either ∞ or 60s. You may choose the evaluation period for the pkts/s as you see fit. Values like 1s, 0.5s, 2s, 5s and 10s are recommended. The evaluation period for both of the traces should be the same.

1. Auto-correlation within one trace
   - Choose one of the traces. Determine the auto-correlation values for packets/s with various values of the lag. Plot the timeseries, and the autocorrelation values into the same plot. Discuss your findings. You can also determine the autocorrelation with different sampling periods (for pkts/s).
2. Cross-correlation between two traces
   - Determine the cross-correlation values with various values of the lag. Plot both of the timeseries, and the cross-correlation values into the same plot. Discuss your findings.

2.3 Correlation in flows

Determine flows/s arrivals for two *different* flow traces. Note, that now you have two new timeseries at your disposal. See the traces-directory in course homepages. You may choose the evaluation period as you see fit. Values like 1s, 0.5s, 2s, 5s and 10s are recommended. The evaluation period for both of the traces should be the same.

1. Auto-correlation within one trace
   - Choose one of the traces. Determine the auto-correlation values for flows/s with various values of the lag. Plot the timeseries, and the autocorrelation values into the same plot. Discuss your findings. You can also determine the autocorrelation with different sampling periods (for flows/s).
2. Cross-correlation between two traces
   - Determine the cross-correlation values with various values of the lag. Plot both of the timeseries, and the cross-correlation values into the same plot. Discuss your findings.

2.4 Comparison

Comment and compare the packet and flow arrival processes based on the stats you determined.

2.5 Extra work

If you have the extra time:

- Determine the *flow* arrival process for one of the dec-files (preferably with the 60 second timeout).
- Do the autocorrelation analysis
- Do cross-correlation analysis with one of the funet-traces.
- Comment on the results

3 Method of working

The reports are written solo. However, a large part of the work that needs to be done might be done quicker and with better results if performed in groups. Therefore, the following is strongly suggested:

- As you begin your work, form groups of three to four students. It would be best if each member of the group would have a different data set to be analyzed to avoid the temptation to copy all the results and to enable the possibility to learn from other data sets.
- Work together as you see fit. If someone in the group is good in making Matlab-scripts let him/her be the one to do the actual coding. If someone else has the ability to quickly form an overall view of the problem and the steps to solve it, use this person's ability. Someone might be more familiar with the typesetting environment you are using. The overall goal is to learn to share your strengths and absorb from other people's work what your own skills lack. However, everyone is required to write their own report and draw up their own conclusions (after discussing about them within the group).
- The final report of each person must contain a section titled "Acknowledgments" where everyone who has influenced the report must be mentioned by name and their individual respective contribution. Your own work should also be clearly defined and reported. It must be stressed that the quality of group work is one key factor of the grading of your work.
- Be advised that more names and more collaboration is not a bad thing. Be also advised that working totally by yourself may lead to rejection of the report. Naturally, plagiarism and 1-to-1 copies of other's reports are considered cheating and lead to appropriate consequences as defined by the Department of Electrical and Communications Engineering and Helsinki University of Technology.
- It is recommended, although not mandatory, to change your working groups during the course. An ideal situation would see you change groups for every exercise.

4 Report structure

- Start with *Introduction*: Introduce the problem, the given input data
- Present the mathematical basics of your solution. Type out the equations and only the equations you need. For instance, do not just tell that you are going to calculate the mean. Instead, state clearly that you are going calculate the mean defined as

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

- Present the complete mathematical solution and process to obtain your results.

- Include the code, command lines and scripts so that anyone can do the same experimentation with their own data and thus verify or contradict your results.
- Present the results in clear figures and tables. Try to avoid complex figures and aim for simplicity. If you know what you are looking for in the data your pictures will usually clearly indicate this. If you're not sure what you're after, your figures and tables will usually reveal this. When you include a figure or a table, be absolutely sure that you refer to it (and explain it) in the text.
- As a final part of your work make conclusions based on the results. If you see causality somewhere, be sure to state it. Be critical and be honest.
- In the Acknowledgements-section you should clearly state all those, by name and contribution, who have contributed and influenced your work. Very few things in the world are accomplished completely alone.
- As a final part of work state all your references and sources of information. The list of references should enable the reader to get their hands on the same material and to verify that you've interpreted the sources correctly.

The report should be as short as possible. Be concise and document all necessary information. However, omit redundancy and avoid unnecessary verbosity. No strict page counts are given. Enough is enough.

5 Grading

The reports are graded either fail, pass or pass with distinction. Reports graded "fail" are rejected and need more work to be passed.

To pass with distinction, the following elements must be present in the report:

- Clear and logical structure in the report.
- Proper language and ease of reading. Note: You are recommended to use English, however, Finnish and Swedish are also accepted. Nevertheless, the get the most out of this course, please use English!
- Clear, and to the point figures. Clear and accurate presentation of the results.
- The problem needs to be clearly defined and properly solved.
- The procedures (scripts, command lines) with which the solutions are obtained need to be clearly documented. The reader of the report has to be able to duplicate the work.
- Clear, concise and to the point conclusions and discussion based on the results are mandatory.
- The work flow and time spent needs to be documented, and other people's contribution clearly identified in the Acknowledgement-section. Also all sources of relevant information (article, program documentation etc.) regarding the solution of the problem need to be presented in the References-section.

5.1 Return of the exercises

Deadline of this exercise is April 9th at 6pm. 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.

6 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.