TECHNISCHE UNIVERSITÄT BERLIN
Fakultät IV – Elektrotechnik und Informatik
Fachgebiet Internet Network Architectures

WirelessLab WS 2016/17
Assignment 3: Tools of the Trade - Data processing and Performance Evaluation

The goal of this assignment is to keep on teaching you the necessary tools to perform and analyze measurements: You learn how to parse a data set and plot the results. Finally, you evaluate the plot, making conclusions about the performance of the links.
The goal of handling, parsing and processing a data set is to make it suitable for analysis. Often, the data is dispersed over different files or needs to be extracted from log files before it can be analyzed. To handle, parse and process such data, you typically use a scripting language. Some people like to use Perl, Python, Ruby or even bash/sh with the combination of awk, cut, find, grep, sed, tr, and xargs commands. *For this course, it does not matter which one you use, what matters is that there is one language that you are comfortable with.*
For the analysis of a processed data set, you typically use programs that implement statistical tools. For instance, you may use Octave, Matlab, R or Python Matplotlib. If you are just getting started with learning to plot, we advise you use R, or Matplotlib, if you are otherwise familiar with Python. In Appendix B, we list a few useful R commands.

Question 1: (100 Points) *Analyzing Wireless Networking Traces*

Download the data set (file `802_11a.tar.bz2`) and look at its structure.

The data set is described in the Appendix A.

(a) For the links 5, 14 and 35, for channels 36, 64 and 165, plot the following:
- A histogram of the RSS with the median and its confidence interval.
  Please make one plot for each combination, so nine plots altogether.
- A moving average (with window length of 16).
  Please make one plot for each combination again and put it next to the histogram.
- A single boxplot that shows all nine combinations next to each other.

Which link is the best, which one is the worst? What else can you see in the plots?

(b) For each combination of link and channel (yes, that means 1014 combinations, 78 links and 13 channels), compute the MAC frame delivery ratio and the median RSS.
Now make a single plot in which each data point corresponds to one combination's median RSS versus the delivery ratio for this combination.
What observations about the performance can you make? Can you explain them?

(c) Plot
- the channel versus all corresponding delivery ratios, i.e., you have the channel number on the x-axis and delivery ratios on the y-axis.
- the link versus all corresponding delivery ratios.
- the channel versus all corresponding RSS.
- the link versus all corresponding RSS.

(1 plot each)
Are there any correlations? Any other observations? Can you explain them?

(d) Consider the data for the links 1 and 17. Plot the ECDFs for all the RSS values of both links (i.e., two lines into a single plot). What do you observe?

Submission
`https://www.isis.tu-berlin.de/2.0/course/view.php?id=8501`
Please submit a PDF document containing *a cover page* with your names and group ID, and *having your group number in its file name.*
The PDF should contain for all questions:

- A description of your approach. Which tools did you use? Where can we find the scripts?
- The plots produced by your scripts. Please number them and reference them in the text, so we know which plot belongs to which question - If we don't understand what you submit, we cannot grade it.
- A description of what you observed from the plots. What data shows the better performance? Are there outliers? Is there any notable phenomenon?
- If applicable: The answers to the discussion questions. What were your expectations about the performance? Does the plot match with your expectations? If not, what could be the explanation?

Please also include your scripts in a format in which we could run them ourselves.
All code must be properly documented using inline comments in English.
Make an archive (.tar.gz, .zip) containing *a directory* with all of your files and *having your group number in its file name.* All files that belong to a specific question must have the question/subquestion in their filenames. Please try not to clutter your submission with temporary files.
Due Date: Wednesday, 16 November at 11.55 p.m. (23:55).

A Description of the data set

The data set contains simplified packet traces of an IEEE 802.11a wireless mesh network. There are 78 links and for each link, data was collected on 13 different channels. The trace shows data frames as they were received, not sent. The trace for each combination of a link-channel was taken in isolation and independently of all the other link-channel combinations.
Each file corresponds to one particular combination of link and channel. Files are in the `csv` format. Each line corresponds to one received frame.
A typical line in any file is

```
954974914,Data,36,23,-73,-96,6.0,3939
```

where 954974914 is a timestamp, `Data` can be ignored, 36 is the channel, 23 is the SNR, -73 is the received signal strength (RSS), -96 is the noise floor, 6.0 is the data rate (fixed for all experiment) and 3939 is a MAC sequence number (between 0 and 4095). This MAC sequence number can be used to compute the delivery ratio.

B Some useful R examples and commands

R command	**Description**
num <- c(1, 2, 3)	stores numbers 1, 2 and 3 to vector "num"
tree <- read.csv(file="tree.csv",head=TRUE,sep=",")	generates a data structure named "tree" from a comma separated CSV file, where the first row is a set of labels(head=TRUE)
median	compute the median
summary	compute the min, max, mean, median, and quantiles
hist	computes and plots a histogram
boxplot	computes and plots a boxplot
ecdf	returns the ECDF of a data set
knots	returns the x values of a step function (e.g. ECDF)

For more information: `http://www.r-project.org/`
For a brief intro: `http://www.cyclismo.org/tutorial/R/`