



3rd Work Sheet Internet Measurement SS 08

Question 1: Network Flow Analysis

In this assignment we will cover the handling of measurement data in form of network flows. A flow is an aggregation of similar packets that are close in time. The most common type of flows, the *Five-Tuple Flows* is the result of aggregating packets with the same source and destination, in the form of IP addresses and port numbers, and by the transport protocol :

(SRCIP, SRCPORT, DSTIP, DSTPORT, PROTOCOL)

Note, that this definition causes a flow to be uni-directional! Thus a TCP connection generates at least two flows, one per direction.

A flow ends, when there has been no matching packet for a certain time called *inactivity timeout* (usually 15 seconds). In the case of TCP, a flow can also be terminated by a FIN or a RST packet.

Flows carry with them information like start end end of a flow, number of bytes and number of packets contained in a flow.

On the lecture's webpage (in section 'Homework') you will find a file in compressed format, `pseudoflows.asc.zip`, that contain flow records generated from real network traffic. Each line contains one flow and the fields are defined as follows:

1. timestamp of first packet
2. timestamp of last packet
3. source node index (IP replaced by unique number)
4. source port number
5. destination node index (IP replaced by unique number)
6. destination port number
7. transport protocol (6=TCP, 17=UDP)
8. payload bytes contained in the flow (no IP/TCP/UDP headers)

Using this flow trace, your task is to

- (a) plot probability density plots and CCDF¹ plots for both the duration and the number of bytes transferred of all flows in the trace.

Compare these plots with the corresponding plots derived from the simulation traces from the second assignment. What differences are visible? What could be the reason for possible differences in the distributions?

- (b) generate density and CCDF plots for duration and size of Web flows and Non-Web flows and compare the results. A Web flow can be classified here as having either source or destination port number 80. Again describe possible differences and what is causing them.

¹Complementary Cumulative Distribution Function

Question 2: Measuring Traffic Variability

In this part you are to generate global scaling plots for the simulation traces from the 2nd assignment. To do this, you first need timeseries of packet counts on the link from node 38 to node 1. A sample perl script to derive such time series is provided on the lecture Web page. For a description of the traces, refer to the 2nd assignment sheet.

In addition, you will need GNU R, a R wavelet package and a R function to generate the plots. *A complete description of how to install and use the wavelet package can be found, along with the scaling plot function on the lecture Web page under*

<http://www.teaching/ss08/IMlecture/#third>

Using the simulation traces and R, you are to do the following:

- (a) Create timeseries for packet counts of both Web and FTP simulation traces at the link from node 38 to node 1 of the simulations of the 2nd assignment. Use the finest time resolution that still makes sense (not too many zeros). Be careful to consider only packets that actually make it to node 1 and don't get dropped due to a filled up output queue at node 38!
- (b) Generate scaling plots for both Web and FTP traffic. Try to adjust the begin and end octaves, j_1 and j_2 , so that you get the closest and widest linear fit of the curve. Write down the resulting Hurst parameters as returned by the plot function.
- (c) Compare the scaling plots and Hurst parameters of Web and FTP traffic. Are there significant differences? If yes, what is the reason for the differences?

Submission Details:

Due Date: June 25th 2008, 10:00 c.t. (just before the lecture) In paper form (i.e., print out the plots; print or write down values like mean). We might also offer "digital submission". If we do, you will find more information on the homepage.

Note: during the next tutorial session (June 25th) we will again present a reference solution of the work sheet.