ELEC-E7130 Network capture

Markus Peuhkuri Tran Thien Thi

Introduction

This exercise will cover the main topics related to capturing and analyzing network traffic using tools such as TCPdump, Wireshark or tshark. After completion of this exercise the students must have good understanding of available techniques for capturing network traffic and how to analyze the captured data. We recommend reading whole document before starting.

In addition to other supporting material, check also *ELEC-E7130 Network capture tutorial* for topics specific for this assignment.

This work contains four tasks:

- Task 1: Analyse flow data
- Task 2: Packet capture
- Task 3: Analyse captured traffic
- Task 4: Compare active and passive measurements

To use some of the course-specific tools, some environment settings are needed. Depending on your login shell you need to run one of following commands in school computer. First command is used if you have any Bourne Shell compatible (like bash in Aalto or zsh).

1. `source /work/courses/unix/T/ELEC/E7130/general/use.sh`
2. `source /work/courses/unix/T/ELEC/E7130/general/use.csh`

You need to provide the tool's name and method (command line if any) you have used to answer above questions in your report file. We recommend that you try to use at *least one command line tool* for analysis because in final assignment the data volume is much larger.

The analysis is most likely two-phase. First to produce initial data and then analyze dataset further to get the answers for some questions.

In addition to the PDF report, you can also provide scripts and other supporting material as zip archive. It will make grading easier in case your report is not fully clear. Do **NOT** include full data files.

Task 1: Analyse flow data

First use a tool (CoralReef, NetMate, tstat or program of your choise) to convert the given sample pcap file (`$TRACE/capture/flow.pcap`) into flows.

I. Provide basic statistics of flow data including - total number of flows, - minumum, median, mean and maximum flow sizes in bytes and packets

II. Which are the top-ten host-pairs based on
   - number of flows
   - number of bytes Are there same pairs?

III. Plot the number of flows for the 100 most common pairs of hosts

1. Using linear scale
2. Using logarithmic scale

IV. Repeat the plot using this time fixed size (2^{16} slots) array approach (solution #2). What can you say about the results?

V. Is there a better way to do this (in terms of running time / memory consumption)?

Note: You can use `/bin/time` command to get resource consumption of a command, use `-v` for more verbose. It provided more detailed output than shell built-in `time`.

Report, task 1

- Describe how you generated flow data
- Provide descriptive statistics
- Provide table of top-ten host pairs
- Provide Top100 plots and evaluate them
- Provide Top100 plots with fixed-array approach.
- Discussion on resource memory requirements.

Task 2: Packet capture

This task must be done with your private computer. If you do this with virtual computer, your of course need to generate traffic within that virtual machine, not at the host. Contact course staff if you do not have a computer to use for this task.

Choose one of the packet capturing tools available to use. The capture takes place in two phases. Make separate capture sessions.

1. Capture network traffic for duration of one hour or more and record interface counters and overall statistics in beginning and end of the packet capture. Use computer normally (do assignments, browse web, check emails, watch video).

2. Again, run a capture for about 15 minutes. In addition to normal use, this time run few iperf3 tests to iperf servers listed in previous assignment. Also run ping toward research servers and iperf servers. Record results of these active measurements. Also record counters at the beginning and at the end.

After you finished capturing the packets try to do the first sanity checks on captured data for

- Size of trace files.
- Number of packets in trace file.
- Total size of packets.
- Compare values from interface counters to capture file. Is there any difference?

Report, task 2

- Describe your measurement setup (tools and workflow).
- Summary of capture data for both sessions.
- Were there differences between capture file statistics and counters?
- Any observations on those two sessions?

Task 3: Analyse captured traffic

Choose one of the mass analysis tools to use (some packet capturing software can also perform analyzing for such small amount of data but it is better idea to practice mass analyzer tool for now). Analyze the **first session** captured data using suitable tool and answer following questions:

I. How many IP (and IPv6 if any) hosts are communicating?
II. How many hosts were tried to contact to, but communication failed for a reason or another? Can you identify different subclasses of failed communications?
III. Top 15 hosts by byte counts.
IV. Top 15 hosts by packet counts.
V. Top 10 TCP and top 5 UDP port numbers (by packet count).
VI. Top 10 fastest TCP connections
VII. Top 10 longest (by time) TCP connections
VIII. Bit and packet rate over time (e.g. **tcpstat**)

Report, task 3

- Describe your analysis setup. Include code sniplets.
- Answers to questions above.
- Did byte and packet count top hosts differ?
- Any interesting observations?

Task 4: Compare active and passive measurements

In this task we compare results we got from active measurements and ones we got from the passive ones. Some helpful guides can be found from supporting material.

At the first you need to extract information of iperf3 sessions. There are two different options. You can use flow tools (Task 1) because most likely each iperf3 run will result an different flow. An another option is to use **tcptrace** to extract information on TCP connections.

For ping results you need to extract ICMP messages from traces, correlate requests to responses and calculate delay and identify possible packet loss.

I. How much there was traffic that was not iperf or ping traffic?
II. Compare iperf results from active and passive measurements. Provide a table.
III. Compare ping results from active and passive measurements. Provide a table.

Report, task 4

- Describe your analysis setup. Include code sniplets.
- Answers to questions above.
- Were there any systematic bias on active and passive measurements?