Computer Networks: Architecture & Protocols (APRC)

Laboratory Assignments — Project 2

- *Passive Network Measurements and Analysis of NetFlow data*

APRC (Fall 2020)
José Legatheaux Martins
Paulo Afonso Lopes

Dep. Informatics
Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa

Project 2: Passive Network Measurements and Analysis of NetFlow data

Computer Networks: Architecture & Protocols

Introduction

Given two files (available for retrieval on the APRC web site) containing information on the flows that crossed (in and out) a network device during a certain period, you are asked to process these files in order to obtain several data on the traffic crossing the devices and, from there, derive some understanding about the network use.

The files are in a format similar to the exports one can get from a NetFlow collector that receives NetFlow packets from a router, as seen in the Lab 5.

The files, and their content, are as follows:

1) **`www.fct.unl.pt.csv`**: contains all the flows that crossed the interface of the www.fct.unl.pt server during one hour in the afternoon of last April 28th. The **`in`** and **`out`** flows have been recorded, for a total of 21362 flows crossing the interface (in and out). Recall that most TCP connections are represented by two flows.

   The www.fct.unl.pt server IP address is 193.136.126.43, which is a public IP address since the machine is reachable from the Internet.

   Most flows start or end in the public Internet and are related to packets exchanged with machines outside the FCT/UNL network, but some flows are directed to a server with a private address inside the FCT/UNL network; that is probably an auxiliary server inside the production network of FCT/UNL.

2) **`bigFlows.csv`**: this is a capture of around 5 minutes of real network traffic on a busy private network's access point to the Internet; it shows many flows and the traffic was originated by different applications. The flows were collected on a router/access point connecting a company's private network (IP prefixes 172.16.0.0 – 172.31.0.0) to the Internet. The flows start or end in a machine with an IP address from that private IP prefix while on the other side of the flow there is a public address.

Assignment & Grading

The assignment has two parts: a mandatory one, graded to a maximum of 140 out of 200 points, and two cumulative optional parts, graded to a maximum of 30 total points each. The completeness and clarity of the report, as well as of the programs source codes are also evaluated.

a) Mandatory part

In the following specification, the term ** internal domain ** means: a) for the computations on the **`www.fct.unl.pt.csv`** file, the IP address of www.fct.unl.pt and the private IP addresses of the FCT/UNL network (range 10.0.0.0/8); b) for the computations on the **`bigFlows.csv`** file, the private IP addresses of the monitored company network (range 172.16.0.0/16 - 172.31.0.0/16).

Also, ** **`in`** ** means traffic / packets / bytes coming from the outside and going into the domain and ** **`out`** ** means traffic / packets / bytes originating in the domain and going to the outside.

Repeat, **for each file**, unless otherwise specified:

a.1) Compute the total number of packets and bytes (in and out) per protocol (TCP, UDP, ICMP, …) contained in the flows.

a.2) Determine the 50 most popular IP addresses external to the domain by number of flows.

a.3) Determine the 50 most popular IP addresses external to the domain by number of bytes.

a.4) Only for the **`bigFlows.csv`** file: What are the top 50 (or less if their variety is smaller) most popular applications used by the computers in the domain? Are you surprised by the result? (i.e., by which applications are most popular). **Note:** There is a Wikipedia page that lists the assigned port numbers, https://en.wikipedia.org/wiki/List_of_TCP_and_UDP_port_numbers. Or you can Google for the IANA's Well Known Port Numbers.

a.5) Aggregate the two flows representing the same TCP connection; count the total number of TCP connections collected, and the total number of TCP connections that started and finished correctly (the ones where flags show that the connection has been opened, used, and finalized by both sides). Explain the several possible technical reasons that justify why there are TCP connections that were not started or ended correctly.

b) Optional parts

b.1) Option 1: Provide two charts representing the average **bit rate per time unit** that crossed the interface or the router **in** and **out** during the collecting period. The resolution of these charts in the horizontal axe (time) should contain at least 60 bars, or values. Thus, if the collection period is 1 hour, each bar represents at most 1 minute. If the collection period is 5 minutes, each bar represents at most 5 seconds.

b.1) Option 2: Consider the two files and represent in a 2D chart the geographic position of the IP addresses outside the domain computed in a.3.

Delivery of the results

You should deliver a zip file containing:

a) the report (in PDF) with your findings, including the results (in tables or other suitable format) concerning the mandatory results; if you addressed the optional parts, this report must also include their results with the appropriate tables and charts (one section per optional part).

b) the source code of the programs you used to compute the results of the mandatory part; and, if you also addressed the optional part(s), the source code of the programs used to compute their results.