Lab 2

Preface

The pcap (packet capture) format is a standard and portable representation of packet-level network traffic. The pcap library (`http://www.tcpdump.org/`) provides many routines[1] to interface with both live (off the network interface) and stored (previously captured) network traffic. You are already familiar with pcap from lab 1; both Wireshark and tcpdump store and read data in *pcap* format.

This lab is designed to familiarize students with the pcap format and library as the basis for performing arbitrarily complex network traffic analysis tasks, especially across large data sets. For example, one might:

- Compute the fraction of web traffic on a link
- Measure the rate of traffic to a particular destination
- Discover anomalous packets
- Find scanning worm traffic
- Etc, etc.

by analyzing packet traces. This lab begins by investigating the pcap format and library/API. We assume basic familiarity with a UNIX environment, including programming tools. As a refresher, a reasonable UNIX tutorial is available at:

`http://www2.ocean.washington.edu/unix.tutorial.html`.

Remember, `man` pages are your friend[2]. If you need help at any point, ask your professors.

Future labs build on the concepts here – it is imperative that you understand these basics in order to be successful with subsequent lab work.

There are 11 questions in the lab. You may collaborate with your classmates, but your solutions and code must be your own. Submit your written lab report, including your program code, online via the CLE site as a single PDF file by October 19, 2018.

[1] `man pcap`
[2] `man man`

1 Getting Started

We recommend using the course Linux VM which contains many useful tools and libraries:

- **Intranet/NPS:** `http://drax.ern.nps.edu/CS4558-v2.ova`
- **Internet:** `https://www.cmand.org/cs4558/CS4558-v2.ova`
- **MD5 Hash:** `ddb185c8ed54ac5f18b745b91c45a85d`

To open the VM after downloading:

- Run VirtualBox on your computer
- Select File —>Import Appliance
  - Choose the `.ova` file you downloaded and wait for the import to complete
- Select CS4558 from the left pane in the main VirtualBox window and click "Start"
- The VM should boot
- Username is: `student`, password is `bad-password`.
- If you have trouble, please ask your professor.

Note that you may perform the lab without this VM, or using a different Linux distribution, or a programming language other than Python (there exist pcap libraries for all major programming languages). However, it will be up to you to install the appropriate tools, libraries, etc. All in-class tutorials will use Python and `dpkt`:
`https://dpkt.readthedocs.io/en/latest/api/api_auto.html#module-dpkt.pcap`

1.1 Telescope Trace

We will be analyzing an anonymized packet trace from a "darknet" or "network telescope."

- Download the compressed packet trace ***inside the VM***:
  Intranet/NPS: `http://drax.ern.nps.edu/telescope.pcap.bz2`
  Internet: `https://www.cmand.org/cs4558/telescope.pcap.bz2`
  MD5 Hash: `8b7d06fe486c3de80b1d9fc475825801`
- Uncompress (this will take a couple minutes, so be patient):
  `$ bzip2 -d <file>`
- Try viewing the trace in either tcpdump or Wireshark.

Clearly, (if it opens at all) the packet trace is much too large for a human to make sense of it, draw conclusions, or understand the traffic statistics. We will write a program using a pcap library to analyze the trace. The appropriate pcap manipulation libraries for Python and C++ are already installed on the VM.

2 Telescope Capture Analysis

2.1 Basic Traffic Stats

Create a Python pcap analysis program with the ability to loop through all of the packets of the trace. Each packet in a pcap trace is preceded by a header as defined in `pcap_pkthdr`. This header contains a UNIX timestamp, among other fields. To iterate through all packets of the trace, you may wish to use, in C, `pcap_loop()` or `pcap_dispatch()` with the appropriate callback (you must create the callback). With python's dpkt, `pcap.Reader` provides a similar iterator.

1. [5 pts] How many packets does the trace contain?
2. [5 pts] What is the average traffic rate (in bits per second) of the trace[3]?

After the per-packet pcap header is the traffic data. Improve your callback to decode IP packets. Your callback should obtain the source and destination IP addresses, the protocol (e.g. TCP, UDP, ICMP, etc), source and destination transport port (where applicable), etc.

Several questions ask for an answer in the form of a distribution. Please note that a distribution *must* represent each value in the form of a fraction or percentage of the total. Answer the following questions:

3. [5 pts] How many IPv4 packets does the trace contain?
4. [5 pts] Estimate the size of the telescope (i.e., how many addresses does it monitor). Provide your estimate in "stroke" notation, e.g., "/8."
5. [10 pts] What is the packet protocol distribution? (A table showing the 5 top protocols and their respective contributions is fine.)
6. [10 pts] Plot a histogram of the packet size distribution.

Note that while pcap is the most popular and widely accepted packet capture format, it has several limitations, many of which we covered in the Paxson lecture. Other packet capture formats exist. For example, "PcapNg," or next-generation pcap, is under active development.

7. [5 pts] Find the documentation for PcapNg online. Briefly (no more than 2 or 3 sentences) describe the differences between pcap and PcapNg.

[3] Be sure to omit the size of the Ethernet (link-layer) header

2.2 Backscatter

We have learned about the types of traffic that arrive at network telescopes, including "backscatter." We can use backscatter arriving at a telescope to identify hosts that are under spoofed-source (D)DoS attack.

8. [5 pts] Explain why the TCP SYN/ACK and TCP RST packets in the trace are likely the result of traffic with spoofed source IP addresses.

9. [5 pts] Besides TCP SYN/ACK and TCP RST packets, what other traffic in the trace is likely the result of spoofed traffic?

10. [10 pts] Identify the host in the trace that experienced the highest-rate DoS attack. Briefly explain how you arrived at your answer.

11. [5 pts] For the host in Q 10, estimate the rate of the DoS attack in packets per second.

2.3 Port Scanning

TCP and UDP packets contain a 16-bit source and destination port. A popular network reconnaissance technique is *port scanning* where an attacker initiates connections to different ports on a host to find those that respond – thereby enumerating the available services.

12. [5 pts] What are the five (5) most frequently probed TCP ports?

To identify port scanners we will examine, for each remote host X, the number of unique destination ports X attempts connections with.

13. [15 pts] Create a cumulative distribution plot of fraction of hosts as a function of the number of unique destination ports. The x-axis is the number of unique TCP ports probed per source, and the y-axis is the cumulative fraction of sources.

14. [5 pts] How many IP sources initiated connections to both TCP port 23 and 2323, but no other ports?

15. [5 pts] Perform some Internet research to determine what is the most likely cause of the TCP port 23 and 2323 traffic in Q 14. Please answer with no more than 2-3 sentences.