NetSec Lab Review Preparation

> Think about it!

Exercise 1

whois

- Think about it. How could somebody on the defense side make the most of whois agains attackers?
- They can find out some information about the attacker from their IP address. Although their personal data would most likely not be registered, some information can usually still be useful (e.g., the use of a certain cloud provider).

IPinfo

- Think about it. Are email servers always hosted by the targeted company? Considering both possibilities, does it make sense to target the email servers?
- Yes: Then the email servers would be a further direct attack vector of the company.
- No: Even if the email servers are hosted externally, targeting them can still affect them, by impacting their communications (DoS) or even by stealing information from the emails if they can be accessed. Of course this is going to be a lot harder if it’s a large email/cloud provider.

rep-1-7

- Passive reconnaissance tools

rep-8

- TCP handshake responses
  - SYN-ACK
  - RST-ACK / RST
  - no response
  - ICMP MESSAGE

- Think about it. Do you consider Wireshark as a suitable tool for analyzing large amounts of network traffic data? Do you think you could detect hosts performing horizontal or vertical scans?

rep9-10

- Brute Force Attack Identification
- Many POST requests from same IP with different username/password combinations

Exercise 2 — Feature Extraction

Using Go-Flows

rep-11

- Finding the 10th packet

rep-12

- Top Protocols
- Think about it. Which advantages does Go-Flows have compared to tcpdump? What are the proportions of TCP, UDP, ICMP traffic
    - go-flows you can configure exactly which features and aggregations you want
    - ? quatsch

rep-13

- AGM flows (AGgregation & Mode Vectors)
- Features are: *srcIP, destIP, srcPort, dstPort, Protocol, TCPflag, TTL* and *pktLength*
- Think about it. Can you think about legitimate and illegitimate situations in which a source sends traffic to many different destinations (hundreds or thousands) in a short time?
    - DOS Attack, Network Scans
    - Some big server responds to many requests. Some Google server

rep-14

- Aggregated statistics
- Think about it. Do the pkts, #uIPdst and bytes correlate
    - pkts and bytes correlate, uipdst don't

Exercise 3 — Time Series Analysis

Internet Background Radiation, or dark-space traffic

consists of IP packets going through the internet but addressing IP destinations that do not exist.

rep-15

- Positive Signal correlation
- Think about it. What could be the reason why the drop in the number of unique IP sources after Jan 16. does not cause a proportional drop in the other signals?
    - They didn't send much

rep-16

- 10x more destinations vs. sources

- Think about it. Do the results make sense for you? Would you expect a different ratio in a normal network (no dark-space)?
    - Maybe not 10x?
- Think about it. You used median in this question, but could have used mean. Does it make a difference? What's better in your opinion? When to use *mean* and when *median*? Can you figure out *pros* and *cons* for both measures of central tendency?
    - Outliers have a smaller impact on the median than on the mean. Pro mean: You don't want to ignore said outliers. Pro median: Ignore one time extraordinary events which would mask "regular traffic".

rep-17

rep-18

- Time series statistics
- Compare same time period in global_last10years.csv to teamX_monthly.csv
- **teamX_Ex3_histograms.png** and **teamX_Ex3_boxplots.png**
- Think about it. Do values in Table A and Table B coincide? If not, why?
    - Outliers only in hourly data: "Averaged away" for daily data.
    - Outliers only in daily data: The effect of slightly outlying hours is only apparent if summed up over the whole day (if it stays that way for several hours/days)
- Histograms and box plots of hourly counts might differ from the equivalent histograms and box plots calculated with daily averaged data. Do you know why? Can you find an explanation?

rep-19

- Think about it. Make sure that you are familiar with the three main protocols appearing in the *teamX_protocol.csv* file. You should know their definition and what are they used for.
    - They are TCP, UDP and ICMP

rep-20

- Residual protocols
- What is the proportion of network traffic that does **not** belong to any of these three most used protocols (let's call it: "others")? Give a percentage with two decimals for (a) packets, (b) number of unique IP sources, and (c) number of unique IP destinations.
    a. 0.03% other packets
    b. -14.86% number of unique sources
    c. -8.14% number of unique IP destinations
- Think about it. Did you get negative values in [rep-20]? Can you figure out why? And why not in the case of packets?
    - Sum of unique IPs for multiple protocols which are not disjoint

rep-21

- Detecting Periodicity
- Think about it. Do signals in rep-21 show periodicity?
    - 744 / 24 = 31, i.e. monthly periodicity

Exercise 4 — Analysis of Flow Records

rep-22

- Think about it. Think about what you would do if you find such artifacts in your data while analyzing network traffic: remove the packet/flow, replace by zero, interpolate? What does it depend on?

rep-23

- Think about it. Top TCP flag values. Does it make sense?
  - 78.77% Syn, 9.77% Ack, 5.03% Push Ack
  - Maybe it makes sense as it’s in the background radiation. Unanswered SYNs?
- Does the TTL plot show mountain-like shapes? If so, can you figure out why?
  - TTL is reduced per hop. Only certain numbers appear in the backbone of the internet.

rep-24

- Bivariate analysis
- Nothing to think about.