School of Computing and Information Systems
The University of Melbourne
COMP90073 Security Analytics, Semester 2 2021
Project 1: Detecting cyberattacks in network traffic data

Release: Tue 3 Aug 2021
Due: 1pm, Tue 24 Aug 2021
Marks: The Project will contribute 15% of your overall mark for the subject. You will be assigned a mark out of 15, according to the criteria below.

Overview

In this project, you are given a network traffic dataset and should use Splunk to identify cyberattacks by leveraging the analytics capabilities of this software. The aim is to strengthen your skills in analysing traffic patterns and identifying their changes over time, which might be signs of suspicious activities. In searching the evidence of cyberattacks, and hunting the attack sources and targets, you will develop the practical security incident investigation skills and mindset of a real-world Cyber Security Analyst. In addition, you will skill-up yourself in tracing attacks back in time to create an attack narrative[1]. Then generating and extracting significant patterns/features of detected attacks will pave the way for the next project that is heavily machine learning focused. Lastly, you will develop your skills as a Cyber Defender by proposing the countermeasures to detect/mitigate similar attacks in the future.

You will write a technical report on your findings, and your proposal on how the identified attack patterns and evidence can be used to detect and mitigate similar cyberattacks in future.

Deliverables

A technical report that describes your methodology for

1. Ingesting the given pcap file into Splunk (1 mark)

   Note: *If you fail to ingest pcap file after multiple attempts, you can ask your Tutor for a copy of the indexed file "<file_name>.pcap.csv". Then copy the file to this directory: "$SPLUNK_HOME/etc/apps/SplunkForPCAP/PCAPcsv/". Please use this as last resort only. Before asking for the indexed file, please be prepared to lose the mark for this deliverable, and you will have to explain what steps you've taken to troubleshoot the issue.*

2. Analyzing the data using Splunk, validating the evidences of the following attack scenarios contained in the given pcap file. You can use either Splunk Search or PCAP Analyzer Dashboard where applicable, new field extraction may be required if you are using Splunk Search.

[1] When the attack was started, the attacker(s), the victim(s) and the type of attack.

2.1 SPAM (2 marks).
   a. Calculate how many email addresses have been targeted by this spam (Hint: search by protocol with the key word of "RCPT")
   b. List the start time and the end time, the first and last recipient (email address) of this email spam (Remember to add the time zone in your answer)

2.2 Port scan (3 marks). TCP SYN scan is one of the most common technique for port scan. Identify the TCP SYN scan activities of 249.56.230.66 in the given dataset, including:
   a. The statistics of packets sent each minute
   b. The number of hosts that were scanned
   c. The start time and the end time

2.3 HTTP (1 mark).
   a. Identify all the URI strings related to the malware downloaded from "my.shopandbuy.com" and "chiashop.net" (Hint: you will need to get the IP address of the web site first)

3. Evaluating the consequences of the attacks on the targeted network (Hint: targeted network is where the infected system belongs to, evaluate the impact using CIA triad) (1 marks)
4. Generating and extracting the significant patterns/features for attack scenarios above, *e.g.,* "*src_IP+src_Port*" can be a significant pattern to detect Flooding DDoS attacks (2 marks)
5. Assuming you are the Cybersecurity Analyst who is part of the Incident Response team, and you've been given the greenlight to put in any controls to mitigate this attack. You can safely ignore any business impact as the priority is to the contain the current attack. Please propose your countermeasures to detect/mitigate the above attacks scenarios, using evidence and patterns in deliverable #***2*** and #***4*** (2 marks)

Technical Report

A technical report of no more than 2500 words in PDF format, comprising:

1. A data description and a summary of detected attacks, including the IP addresses of attackers and victims, the attacked services, the timestamp, and the type of the attack per attack scenario.
2. Methodology of analysis to find evidence of cyberattacks in the network traffic data.
3. Description of each attack and the attack narrative.
4. Possible approaches for extracting features (fields) and summary of your approach.
5. Proposed countermeasures per attack scenarios.
6. Conclusions

You should include a bibliography and citations to relevant research papers and external resources and codes you have used (these will not be count in the word limit).

Assessment Criteria

Report (15 marks out of 15)

1. Methodology: (7 marks)
   You will describe your methodology in a manner that would make your work reproducible. You should describe in detail how you have detected the cyberattacks using Splunk search capabilities: the exact SPL commands you ran and the corresponding generated results, or the dashboards you used in pcap app (*PCAP Analyzer)* and the corresponding data and generated results. Your approach to model patterns in data and detect changes in them for identifying cyberattacks should be clearly explained. The description of your proposed countermeasures should include reasons for choosing it based on the types of attacks you have detected. You should not use a network traffic feature generator as a black box without explaining the reasons for extracting the reported features.
2. Critical Analysis: (5 marks)
   You should validate the evidence you have found in the data for proving that a certain type of attack has happened. The attack narrative should specify the time of start and end of the attack and the consequences of the attack on the victim network.[1] You should identify other types of data that could be collected to more accurately and effectively detect/mitigate the identified cyberattacks.
3. Report Quality: (3 marks)
   You will produce a formal report and express your methodology and findings concisely and clearly. The quality and description of figures and tables should be acceptable. In real-world scenarios, this report will have a range of audience in a company. Thus, it should be structured such that summary of the findings is available for the managers and non-technical audience and on the other hand, the attack narratives should include technical details for other analysers that may read your report.

Description of the Data

The dataset for Project 1 (download link) includes packet capture (pcap) file of network traffic for a network that was victim of cyberattacks. This file was captured on the interfaces of virtual machines being infected by malware, which contains attacks including but not limited to three scenarios listed in the Deliverable section. It also contains normal traffic prior to malware infection. This enables you to compare the patterns in data from normal operation (before infection) and post infection to identify attacks that occurred.

Changes/Updates to the Project Specifications

If we require any changes or clarifications to the project specifications, they will be posted on the LMS. Any addendums will supersede information included in this document.

[1] We validate your findings with our ground truth, and you lose marks for identifying non-anomalous traffic as anomaly and vice-versa.

Academic Misconduct

For most people, collaboration will form a natural part of the undertaking of this project. However, it is still an individual task, and so reuse of ideas or excessive influence in algorithm choice and development will be considered cheating. We will be checking submissions for originality and will invoke the University's Academic Misconduct policy (http://academichonesty.unimelb.edu.au/policy.html) where inappropriate levels of collusion or plagiarism are deemed to have taken place.

Late Submission Policy

You are strongly encouraged to submit by the time and date specified above, however, if circumstances do not permit this, then the marks will be adjusted as follows. Each day (or part thereof) that this project is submitted after the due date (and time) specified above, 10% will be deducted from the marks available, up until 5 days has passed, after which regular submissions will no longer be accepted.

Extensions

If you require an extension, please email Yujing (yujing.jiang@unimelb.edu.au) using the subject 'COMP90073 Extension Request' at the earliest possible opportunity. We will then assess whether an extension is appropriate. If you have a medical reason for your request, you will be asked to provide a medical certificate. Requests for extensions on medical grounds received after the deadline may be declined. Note that computer systems are often heavily loaded near project deadlines, and unexpected network or system downtime can occur. Generally, system downtime or failure will not be considered as grounds for an extension. You should plan ahead to avoid leaving things to the last minute, when unexpected problems may occur.

