Swinburne University of Technology

Faculty of Science, Engineering and Technology

ASSIGNMENT AND PROJECT COVER SHEET

Unit Code: COS3005 Unit Title: IT Security

Assignment number and title: Assignment 3 (Research Report) Due date: 31/10

Lab group: _____ Tutor: Yasas Akurudda Liyanage Don Lecturer: _____

Family name: Minh Hoang Identity no: 104487115

Other names: Duong

To be completed if this is an INDIVIDUAL ASSIGNMENT

I declare that this assignment is my individual work. I have not worked collaboratively, nor have I copied from any other student's work or from any other source except where due acknowledgment is made explicitly in the text, nor has any part been written for me by another person.

Signature: 

To be completed if this is a GROUP ASSIGNMENT

We declare that this is a group assignment and that no part of this submission has been copied from any other student's work or from any other source except where due acknowledgment is made explicitly in the text, nor has any part been written for us by another person.

ID Number	Name	Signature

Marker's comments:

Total Mark: _____

Extension certification:

This assignment has been given an extension and is now due on _____

Signature of Convener: _____ Date: / 2024

COS3005 – IT Security

Research Report (Assignment 3)

Topic: *Incident Forensic Analysis (Review of Literature)*

Student: *Minh Hoang Duong (104487115)*

Student Name: *Minh Hoang Duong*

Student ID: *104487115*

Word Count: *(excluding reference, coversheet, title page, and multiple titles)*

Due Date: *31/11/2024*

Submission Date: *31/11/2024*

A. Introduction

Throughout the development of the digital age, cybersecurity has always played an important role in protecting sensitive information and users against many cyber threats. These threats have become increasingly sophisticated, urging corporations and companies to adopt up-to-date and effective detection systems and mitigation strategies. This report aims to review the current trends and current cyber security incidents in threat detection and mitigation strategies, focusing on the implementation of advanced and trending technologies such as machine learning, and statistical methods. In addition, understanding data features being used to help with detection and mitigation is crucial in improving and enhancing detection capabilities. This report aims to highlight good practices and trends in combating cyber threats by reviewing and analyzing reputable literature.

B. Approached to current trends

What are the current trends to detecting and mitigating cyber threats like or closely related?

Currently, as technology evolves in various fields has led to a growing number and sophisticated cyber-attacks. Machine Learning and Artificial Intelligence have been recognized as effective tools in combating these threats as there has been a growing interest in the uses of Artificial intelligence (AI) and Machine Learning (ML) cyber security with 45% [6] of organizations already implemented and 35% [6] plan to follow [6].

- One of the notable trends is the implementation of **Machine Learning** to enhance suspicious activity detection. According to the IEEE Journal [2], Multiple machine learning/Deep learning models play a vital role in threat detection and response such as spam classification, phishing, fraud detection, malware detection, and intrusion detection [2]. Specifically, commonly used classification Machine Learning models like J48, Bayes Net, SVM, Decision Tree, Random Forest,… have been used for phishing emails, and phishing website detection [2]. These models used pre-labeled data for training in pattern recognition, therefore accurately predicting the outcome of new input [4]. More specifically, they distinguish between legitimate and phishing emails/Websites or URLs, and websites (which is a classification problem).
- In this report, published at the 2022 7th International Conference for Convergence in Technology (I2CT), hosted in Mumbai, India [3], the researchers attempted to differentiate between legitimate and phishing emails by applying a dataset containing 1000 URLs in which 5000 are phishing and 5000 are legitimate to 2 classification machine learning models, including Decision Tree and Random Forest. After training, the accuracy score of the random forest is 87.0%[3] while the accuracy score of the Decision Tree is 82.4%[3], proving the potential of Machine Learning models in phishing detection and prevention.
- In other reports, published at the 2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT), hosted in Alkhobar, Saudi Arabia [5], researchers evaluated eleven commonly used Machine Learning Algorithm including Decision Tree[5], K -Nearest Neighbour (KNN)[5], Gradient Boosting (GB)[5], Logistic Regression (LR)[5], Naïve Bayes (NB)[5], Random Forest (RF)[5], Support Vector Machines (SVM)[5], Neural Network (NN)[5], Ex-tra_Tree (ET)[5], Ada_Boost (AB)[5] and Bagging (B)[5] with the application of pre-collected URLs including 2843 legitimate URLs and 8495 phishing URLs. The result shows that extra-tree (ET) and K Nearest Neighbour had the best accuracy of 91% proving their potential to distinguish between phishing and legitimate URLs.

Another cyber threat detection and mitigation approach that has been growing currently is **Honeypot**. Since its first implementation in 2003 [9] there has been an increase in Honeypot research recently [7] with the majority of publications coming from the USA followed by South Africa[7].

- One notable publication by outpost24[8] highlighted the role of honeypot deployment in threat analysis. This research outlines the implementation of how honeypot deployment serves as an effective tool for gathering actionable information about active cyber threats. By distributing multiple types of honeypot systems around the world, more specifically 3 in North America, 6 in Asia, 8 in Europe, and 3 in Oceania, the publication reported finding and analyzing over 42 million registered attacks [8] toward the system, proving the effectiveness of honeypot in studying active threats.

- Furthermore, the report published at the 2020 8th International Conference on Cyber and IT Service Management (CITSM), hosted at Pangkal, Indonesia, highlighted the growing interest in researching publications for honeypot implementation for Machine Learning Malware Detection. The publication revealed that this research trend fluctuated from 2012 to 2020, with a significant improvement in 2019, favoring honeypot virtualization techniques [10]. The primary advantage of this approach lies in the ability of Honeypot to gather intelligence from active cyber threats. Once the data is collected, it can be processed and used to apply to training multiple Machine Learning models such as those listed above.

C.The use of features

What data features are being used to aid detection or mitigation?

In the report published at the 2021 4th International Symposium on Advanced Electrical and Communication Technologies (ISAECT) mentioned above [5], the researchers attempted to apply lexical analysis to extract and analyze URL features including the **number of '.'** (Dot_count), **number of characters** (url_len), **number of digits** (Digit_count), **if the https protocol is used** (Protocol), **number of "-"** (Hypen_count), **number of "//"** (Double_slash_count), **number of "/"** (Single_slash_count), **Number of special characters (";", ":", "#", "!", "%", "~", "+", "_", "?", "=", "&", ",")** (Special_characters_count), **Number of "@"** (At_sign_count). The features are then applied to pattern-recognition training utilizing 11 previously mentioned Machine Learning models.

On the other hand, the published report on the 2022 7th International Conference for Convergence in Technology (I2CT) mentioned above [3], the researchers take an initiative approach by analyzing 17 technical URL features including **Have_IP, Have_at, URL_length, URL_Depth, Redirection, HTTPS_domain, TinyURL, Prefix/Suffix, DNS_Record, Web_Traffic, Domain_Age, Domain_End, iFrame, Move_over, Right_click, Web_forwards, Label**, without further explanation on each features themselves. The features are then applied to model training utilizing 11 previously mentioned Machine Learning models. The features are then applied to model training utilizing 2 previously mentioned Machine Learning models.

On the honeypot side of the trends, a research journal published at the 2019 3rd International Conference on Trends in Electrics and Informatic (ICOE), hosted at Tirunelveli, India, proposed the idea of applying Machine Learning Detection Framework on Honeypot for defending against DDoS attacks. By capturing traffic including **source IP addresses, source ports, destination IP addresses, destination ports, packet size,** and **timestamp of all sent IP packets**, the honeypot system can effectively group packets based on the source IP address. This grouping enables the system to identify behaviors associated with each unique source. After multiple feature extraction and engineering described in the report, the processed received data are finally fed into the classification Machine Learning model (not specified, but some are mentioned including K-nearest neighbors (KNN), random forests, support vector machines (SVM)) to distinguished between normal traffic and malicious DDoS traffic flow.

D.Discussion

Throughout the literature review displayed above, the application of Machine Learning and honeypot systems in enhancing threat detection and prevention has proven very promising. As indicated above, regular uses Machine Learning models such as Decision Tree, KNN, Random Forest, and Extra-Tree (ET) are capable of handling cybersecurity classification tasks such as distinguishing between legitimate and phishing URLs, normal traffic, and malicious DDoS traffic. In addition, honeypot systems contribute towards intelligence gathering aspect for both multiple active threats, anomaly analyzing, and Machine Learning model training. However, these techniques also come with many limitations. Traditional supervised Machine Learning models such as Decision Trees, Random Forests,.etc relies heavily on carefully prepared, specific data features and input, limiting specific tasks and data types. For example, ML-based for phishing URL detection may perform well in detecting phishing URLs but cannot be implemented in email phishing detection as they require different input and training features. Another limitation lies within the preprocessing and feature extraction stage before being implemented in the ML model itself. For example, the URL before being fed in the ML model for phishing prediction, must be broken down and analyzed for number of dots, protocols,… as the model itself has been trained with those features. Additionally, honeypot data requires careful cleaning and

consideration before being fed into the model. Even though these stages can be automated, their strict nature can still raise annoyance in the implementation stages. On the other hand, emerging fields of Machine Learning such as Deep Learning can address those drawbacks, by offering less reliance on pre-processed, analyzed, and cleaned features. In conclusion, while Machine Learning and Honeypot in cyber security offer powerful detection capabilities, they also come with multiple drawbacks that need to be and could be addressed.

E. Reference

1. Five tech trends for 2024. (2023, August 15). University of Technology Sydney. https://www.uts.edu.au/about/faculty-engineering-and-information-technology/postgraduate/articles/five-tech-trends-2024
2. A survey on machine learning techniques for cyber security in the last decade. (2020). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9277523
3. Phishing website detection using machine learning. (2022b, April 7). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9824801
4. What is Supervised Learning? | Google Cloud. (n.d.). Google Cloud. https://cloud.google.com/discover/what-is-supervised-learning
5. Machine Learning Algorithms Evaluation for phishing URLs Classification. (2021, December 6). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9668489
6. Mohamed, N. (2023). Current trends in AI and ML for cybersecurity: A state-of-the-art survey. Cogent Engineering, 10(2). https://doi.org/10.1080/23311916.2023.2272358
7. A survey of honeypot research: Trends and opportunities. (2015, December 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/7412090
8. Outpost. (2024, April 18). Cyber Threat Landscape Study 2023: Outpost24's honeypot findings from over 42 million attacks. Outpost24. https://outpost24.com/blog/honeypot-findings-from-over-42-million-attacks/
9. Niclas Ilg, Paul Duplys, Dominik Sisejkovic, & Michael Menth. (n.d.). A survey of contemporary open-source honeypots, frameworks, and tools. In ScienceDirect. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S108480452300156X
10. The Use of Honeypot in Machine Learning Based on Malware Detection: A Review. (2020). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9268794
11. A Honeypot with Machine Learning based Detection Framework for defending IoT based Botnet DDoS Attacks. (2019, April 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8862720