Literature Review

## 2.0 Introduction

This chapter consists of related literature on phishing detection and prevention mechanisms acquired from books, journals, academic papers and articles. This includes research on related approaches to phishing detection, machine learning algorithms that have been used to detect phishing, browser security indicators and features that can be used to flag websites as legitimate or fraudulent.

## 2.1 Related Works

Research on phishing detection and prevention mainly explores four areas; automating phishing detection, providing user interface cues to help users detect phishing, educating users about protecting themselves and understanding user's vulnerability (Alsharnouby, Mohamed & Alaca, Furkan & Chiasson, Sonia. 2015). This research mainly explores automating phishing detection and providing user interface cues to help users to detect phishing.

## 2.2. Automated phishing detection

Automatic phishing detectors exist at various levels such as web browser tools, internet service providers and mail servers and clients (Alsharnouby et. al. 2015). These tools restrict access to detected websites or request the website's internet service provider to take it down. Automatic email classification tools make use of machine learning algorithms, spam filter techniques and statistical classifiers to identify potential phishing messages. They have a varying degree of effectiveness and misclassifications are a common occurrence which affects the reliability of the service as users are likely to be intolerant to loosing legitimate messages.

Automated phishing detection techniques to detect phishing websites include the use of blacklists and whitelists, the use of heuristic methods, and use of machine learning principles. (Anjum & Shabut et al., 2016)

## 2.2.1 Blacklist and whitelists Approach

This method often maintains a list of URLs that are labelled as malicious or benign. Whenever a new URL is visited, a database lookup is performed. The URL is checked whether it exists on the list and if it is found the label allocated to it is returned as output. A major problem with this method is the inability to maintain a list of all possible malicious URLs as new URLs can be easily generated daily, thus making it impossible for them to detect new threats. This is a critical concern when the attackers generate new URLs using algorithms, and can therefore bypass all blacklists as the URLs are dynamic (Doyen, Chenghao & Steven, 2019)

PhishTank is a popular blacklist launched in 2006 and has been in service ever since (LLC OpenDNS. 2016). The blacklist is populated through crowdsourcing volunteers who submit potential phishing websites and vote on the legitimacy of websites. PhishTank is not protection. "PhishTank is an information clearinghouse, which helps to pour sunshine on some of the dark alleys of the Internet. PhishTank provides accurate, actionable information to anyone trying to identify bad actors, whether for themselves or for others" (LLC OpenDNS, 2016). PhishTank offers a blacklist for use by other tools through an API. Popular organizations such as Kaspersky, MacAfee, APWG and Avira make use of Phishtank's dataset.

## 2.2.2 Heuristic Approach

Heuristic approaches are similar to blacklist methods since their basic idea is to create a blacklist of signatures. When common attacks are detected, a signature is assigned to the type of attack. The idea is to look out for a signature of malicious activity such as unusual process creation, repeated redirection etc. Intrusion detection systems are able to detect these behaviors and respond to them appropriately. These approaches are able to detect new threats but to a limited extent since new threats may be completely unrelated. Modern heuristic methods analyze the execution dynamics of webpages. They require visiting the actual URL which may initiate the attack. The techniques are resource intensive and require complete execution of the code (including the server-side scripts). The techniques may go undetected since the malware in place may not launch the attack immediately (Sahoo, Liu & Hoi, 2019).

## 2.2.3 Machine Learning Approach

These are intelligent heuristic-based methods which try to analyze the information of a URL and its corresponding webpages by extracting the features of URLs and training a prediction model using training data. In static analysis, the website is analyzed based on the features extracted from the URL string such as lexical features, information about the host, and sometimes the HTML and JavaScript content. The underlying assumption is that the distribution of these features is different for malicious and benign URLs. Using this distribution information, a prediction model can be built, which can make predictions on new URLs. Since no execution is required, they are safer than dynamic approaches which require complete execution of the URL. Dynamic methods monitor the behavior of the websites looking for anomalies. (Sahoo et. al., 2019)

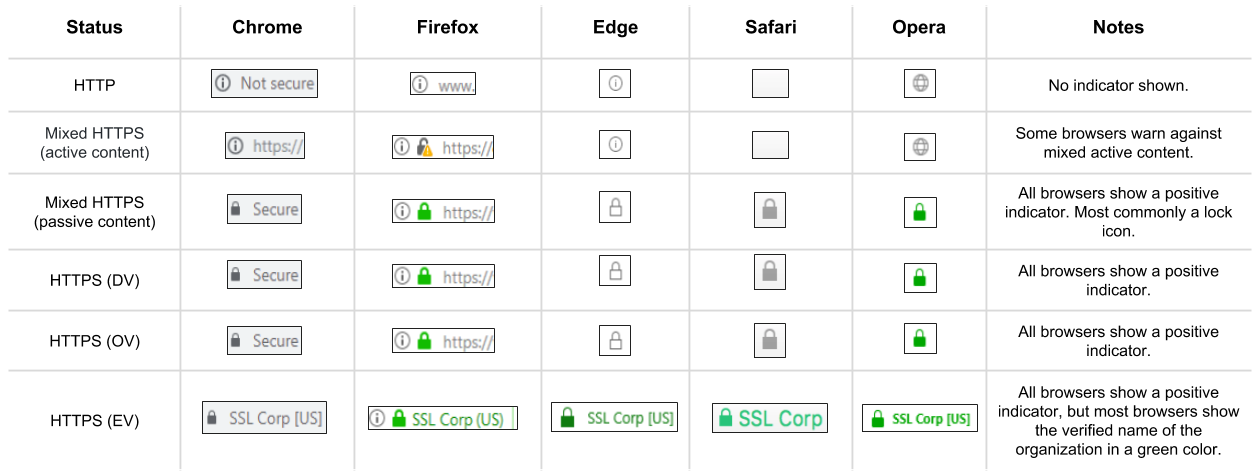
## 2.3. Security indicators

Phishing is primarily a problem because users are unable to verify the authenticity of a website. Security indicators are readily available every time a user launches the web browser. However, an average internet user rarely pays attention to them or may not understand their meaning. In general, most authors consider URL bar HTTPS indicators (SSL/TLS) and digital certiﬁcate indicators as the main cues (Jelovčan, Luka, Vrhovec, Simon, Mihelič, Anže. 2020)

Hypertext Transmission Protocol (HTTP), the most widely used protocol in the world, is the protocol that is used to view web pages. When you type an address, say www.example.com, HTTP is added automatically at the beginning of the address, ie http://www.example.com. HTTP sends and retrieves all data in clear text making it vulnerable to anyone who wants it, eg hackers. Secure Hypertext Transmission Protocol (HTTPS) is used to secure communications by encrypting the data exchanged between a person’s browser and the web site he or she is visiting. HTTPS is especially important on sites that offer online sales or password-protected accounts. Browsers indicate that a website uses HTTPS by use of a padlock (colored in some browsers). (Jelovčan et. al., 2020 & APWG 2021)

Secure Hypertext Transmission Protocol (HTTPS) uses Secure Sockets Layer (SSL), a protocol used to ensure security on the internet using public key encryption. When a computer connects to a website using SSL the computer's browser will ask the website to identify itself. The web server responds by sending the computer a copy of its SSL certificate, a small digital certificate used to authenticate the identity of a website. Once the browser establishes trust with the website, encrypted data can be transmitted to and from the website. (Jelovčan et. al., 2020)

The following table is a summary of the general state of security indicators in modern browsers. Starting with HTTP (which is not secure at all) each item further along the list is more secure than the previous ones. (Naziridis, 2018)



To help users stay safe on the internet, browsers require websites to use certificates from trusted organizations. This is because anyone can create a certificate (e.g., Using OpenSSL) claiming to be whatever website they claim to be. According to APWG (2021), PhishLabs, an active contributor to APWG, has been tracking the proportion of phishing sites that are protected by the HTTPS encryption protocol. Studying HTTP on phishing sites provides insight onto how phishers are fooling Internet users by turning an Internet security feature against them. 83% of the phishing attacks reported to APWG in the first quarter of 2019 used HTTPS protocol.

## 2.4 Machine Learning Algorithms

Several machine learning algorithms have been implemented in real life applications. (Ali, Waleed, 2017). The identification of malicious and legitimate URLs can be regarded as a classification task and some of the popular machine learning algorithms applicable include the following:

1. **Back-Propagation Neural Network (BPNN)**

BPNN are popular algorithms in network models. They are particularly used in prediction and classification problems. They learn in two phases: the forward pass and the backward pass. In the forward pass, the input layer is presented with a training input pattern which is propagated from layer to layer until the desired output is produced. In the backward phase, the output is compared with the anticipated output in order to compute the error. The error is then propagated backward through the network from output to input layers and the weights are adjusted accordingly to minimize the error (Ali et. al., 2017).

1. **Radial Basis Function Network (RBFN)**

RBFN is a type of neural network that uses radial basis functions as activation functions. In the architecture of RBFN, there are three feedback networks: the input layer, the hidden layer and the output layer. In each hidden unit, a radial activation function is implemented while a weighted sum of outputs of hidden units is implemented for each output unit. Learning is conducted in two phases. The first stage involves clustering in order to determine the centers and widths of the hidden layer. In the next phase, the weights connecting the hidden layer with the output layer are optimized through the use of Least Mean Squared (LMS) or Singular Value Decomposition (SVD) algorithms. (Ali et. al., 2017).

1. **Support Vector Machine (SVM)**

SVM, very popular and robust machine earning techniques have been utilized effectively in many applications. They are based on maximizing the margin and thereby creating the largest possible distance between the hyperplane and the instances in order to reduce an upper bound on the anticipated generalization error. Support vectors close to the hyperplane provide the most useful information for classification. An appropriate kernel function is used to transform the data into a high-dimension to use linear discriminate functions (Ali et. al., 2017).

1. **Decision Tree and Random Forest (RF)\* RF ONLY**

In decision trees, a node corresponds to a feature of an instance being classified. The instances are classified through sorting based on feature values. Each branch represents a value that the node can predict. Random Forest is a popular decision tree that can be used for classification and regression. RF is a group of decision trees trained independently on selected training datasets. The classification is then determined by voting among all the trained decision trees. (Ali et. al., 2017).

The performances in terms of correct classification rate (CCR) of the above algorithms were compared together in a study (Ali et. al., 2017). The following table summarizes the results.

|  |  |
| --- | --- |
| **Classifier** | **Correct Classification Rate** |
| Back-Propagation Neural Network (BPNN) | 0.970 |
| Radial Basis Function Network (RBFN) | 0.928 |
| Support Vector Machine (SVM) | 0.963 |
| Random Forest (RF) | 0.971 |

*Fig: Performance measures of the machine learning classifiers*

The Random Forest and Back-Propagation Neural Network classifiers achieved the best correct classification rate while the Radial Basis Function Network attained the lowest.

## 3. Features Extraction

Several features can be extracted from a website to distinguish phishing websites from legitimate ones. Feature selection is necessary in order to decrease computation time and to reduce noise and irrelevant features. The choice of extracted features is critical for the success of the detection mechanism in place. Once the selected features are selected, the machine learning model can be trained. (Ali et. al., 2017)

The following features can contribute to the effective prediction of the phishing websites: (Ali et. al., 2017).

|  |  |
| --- | --- |
| Feature Category | Feature Name |
| Address bar-based features | Using the IP Address  Long URL to Hide the Suspicious Part  Using URL Shortening Services “TinyURL”  URL’s having “@” Symbol  Redirecting using “//”  Adding Prefix or Suffix Separated by (-) to the Domain  Sub Domain and Multi Sub Domains  HTTPS (Hyper Text Transfer Protocol with Secure Sockets Layer)  Domain Registration Length  Favicon  Using Non-Standard Ports  The Existence of “HTTPS” Token in the Domain Part of the URL |
| Abnormal-based features | URL of Anchor  Links in <Meta>, <Script> and <Link> tags  Server Form Handler (SFH)  Submitting Information to Email  Abnormal URL |
| HTML and JavaScript-based features | Website Forwarding  Status Bar Customization  Disabling Right Click  Using Pop-up Window  IFrame Redirection |
| Domain-based features | Age of Domain  DNS Record  Website Traffic  Page Rank  Google Index Number of Links  Pointing to Page  Statistical-Reports Based Feature |

## 4. Why Phishing Still works

According to Retruster 2019, the problem lies in the detection and reporting of cybercrimes. It can take as long as 50 days from when a breach is discovered until the time when it is reported, a very huge risk for potential victims.

Users consider security as a secondary task. They are prone to concentrating on the real purpose of their interaction with their website making it unlikely for them to notice the security indicators displayed. Some security indicators are also only visible when visiting safe and secure websites (Alsharnouby, et. al., 2015). In a study conducted to assess whether browser security indicators and increased user awareness on phishing have led to users’ improved ability to protect themselves from phishing, a series of websites was presented to participants and they were asked to identify phishing websites. Participants were successfully able to detect only 53% of phishing websites even when forewarned to identify them. Using eye tracking, they found that two thirds of users looked at the SSL lock icon when prompted to be security-conscious but rarely used other cues on the browser-chrome (Alsharnouby, et. al., 2015). Users spend 85% of their time looking at the website content during a web interaction and only 6% of their time looking at security indicators. Even if users doubt the authenticity of the websites, they will still access it, primarily because they want the beneﬁts from it (Jelovčan et. al., 2020).

One major problem with cybercrime is establishing jurisdiction. Physical crimes are bound to a physical location. The crime is therefore considered territorial and its location determines the jurisdiction. Cybercrime activities are not bound to a physical location as the victim and the perpetrator can even be physically at different countries. To worsen the situation, some countries may not have extradition treaties and the law may therefore not be able to prosecute the perpetrators. Legislation of laws on cybercrime is still developing, and it may be challenging to bring justice to those who commit cybercrimes (Plachkinova, 2021).

For these reasons, developing a usable browser security to alert users when they are in danger remains a crucial and unsolved problem in security.