WEB PHISHING DETECTION

LITERATURE SURVEY

1. **Min Wu1, Robert C. Miller, Simson L. Garfinkel (2006)**

Security toolbars in a web browser show security-related information about a website to help users detect phishing attacks. Because the toolbars are designed for humans to use, they should be evaluated for usability that is, whether these toolbars really prevent users from being tricked into providing personal information. We conducted two user studies of three security toolbars and other browser security indicators and found them all ineffective at preventing phishing attacks. Even though subjects were asked to pay attention to the toolbar, many failed to look at it; others disregarded or explained away the toolbars' warnings if the content of web pages looked legitimate. We found that many subjects do not understand phishing attacks or realize how sophisticated such attacks can.

2. **Yue Zhang, Jason Hong, Lorrie Faith Cranor (2007)**

Phishing is a significant problem involving fraudulent email and web sites that trick unsuspecting users into revealing private information. In this paper, we present the design, implementation, and evaluation of CANTINA, a novel, content-based approach to detecting phishing web sites, based on the TF-IDF information retrieval algorithm. We also discuss the design and evaluation of several heuristics we developed to reduce false positives. Our experiments show that CANTINA is good at detecting phishing sites, correctly labeling approximately 95% of phishing sites.

3. **Jason Hong (2010)**

Phishing attacks are a significant threat to users of the Internet, causing tremendous economic loss every year. In combating phish, industry relies heavily on manual verification to achieve a low false positive rate, which, however, tends to be slow in responding to the huge volume of unique phishing URLs created by toolkits. Our goal here is to combine the best aspects of human verified blacklists and heuristic-based methods, ie, the low false positive rate of the former and the broad and fast coverage of the latter.

4. **Jun Ho Huh (2012)**

We propose a new phishing detection heuristic based on the search results returned from popular web search engines such as Google, Bing and Yahoo. The full URL of a website a user intends to access is used as the search string, and the number of results returned and ranking of the website are used for classification. Most of the time, legitimate websites get back large number of results and are ranked first, whereas phishing websites get back no result and/or are not ranked at all.

5. **International Journal of Computer Science and Engineering Survey (IJCSES) (2015)**

Phishing is the fraudulent acquisition of personal information like username, password, credit card information, etc. by tricking an individual into believing that the attacker is a trustworthy entity. It is affecting all the major sector of industry day by day with lots of misuse of user's credentials. So in today online environment we need to protect the data from phishing and safeguard our information, which can be done through anti-phishing tools. Currently there are many freely available anti-phishing browser extensions tools that warns user when they are browsing a suspected phishing site. In this paper we did a literature survey of some of the commonly and popularly used anti-phishing browser extensions by reviewing the existing anti-phishing techniques along with their merits and demerits.

6. **Hiba Zuhair**

Phishing has become a substantial threat for internet users and a major cause of financial losses. In these attacks the cybercriminals carry out user credential information and users can fall victim. The current solution against phishing attacks are not sufficient to detect and work against novel phishes. This paper presents a systematic review of the previous and current research waves done on Internet phishing mitigation in different areas of expertise and highlighted phishing attacks types and some existing anti-phishing approaches. Further the discussion about novel phishes and identify the elements of issues highlighted. The review can be valuable source of information to find and identify recent gap and challenges to fulfil the security flaws.

7. **Azeez Nureni Ayofe**

Phishing is a form of social engineering or website forgery whereby attackers mimic a trusted website or public organization or sending e-mails in an automated manner in order to steal sensitive information or credentials of online users. This is done in a way the user does not realize he is in a phishing environment and in turn reveals his sensitive information such as credit card information, employment details, online shopping account passwords and bank information. Phishers are still having their ways to succeed in their various nefarious activities and attacks. Different anti-phishing schemes however have emerged but phishers still find their ways around by breaking through various existing techniques. Against this backdrop, this project aims at developing a web enabled anti-phishing.

8. **Nazrul Hoque1, Dhruba K. Bhattacharyya1, Jugal Kalita (2015)**

Threats of distributed denial of service (DDoS) attacks have been increasing day-by-day due to rapid development of computer networks and associated infrastructure, and millions of software applications, large and small, addressing all varieties of tasks. Botnets pose a major threat to network security as they are widely used for many Internet crimes such as DDoS attacks, identity theft, email spamming, and click fraud. Botnet based DDoS attacks are catastrophic to the victim network as they can exhaust both network bandwidth and resources of the victim machine. This survey presents a comprehensive overview of DDoS attacks, their causes, types with a taxonomy, and technical details of various attack launching tools. Furthermore, a list of important issues and research challenges is also reported.

9. Joshua Saxe, Konstantin Berlin (2017)

For years security machine learning research has promised to obviate the need for signature based detection by automatically learning to detect indicators of attack. Unfortunately, this vision hasn't come to fruition: in fact, developing and maintaining today's security machine learning systems can require engineering resources that are comparable to that of signature-based detection systems, due in part to the need to develop and continuously tune the "features" these machine learning systems look at as attacks evolve. Deep learning, a subfield of machine learning, promises to change this by operating on raw input signals and automating the process of feature design and extraction. In this paper we propose the expose neural network, which uses a deep learning approach we have developed to take generic, raw short character strings as input, and learns to simultaneously extract features and classify using character-level embeddings and convolutional neural network., expose outperforms manual feature extraction based baselines on all of the intrusion detection problems we tested it on, yielding a 5%-10% detection rate gain at 0.1% false positive rate compared to these baselines.

10. Mehmet Korkmaz

A machine- learning based phishing detection system by using 8 different algorithms on three different datasets. The algorithms used were Logistic Regression (LR), K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Decision Tree (DT), Naive Bayes (NB), XGBoost, Random Forest (RF) and Artificial Neural Network (ANN). It was observed that the models using LR, SVM and NB have low accuracy rate. In terms of training time, NB, DT, LR and ANN algorithms gave better results. They concluded that RF algorithm or ANN algorithm may be used because of less training time along with a high accuracy rate.