Literature Survey - Web Phishing Detection

Team Members:

Muazzam N Alseri [311019104051]

Mohamed Suhaib Ahmed [311019104050]

Dilip Kumar K [311019104031]

Kishore G [311019104040]

S.No:	Paper Title	Paper Authors	Published Month and Year	Abstract	Drawbacks
1.	PHISHING DETECTION – A LITERATURE SURVEY	Mahmoud Khonji Youssef Iraqi Andy Jones	April 2013	This article surveys the literature on the detection of phishing attacks. Phishing attacks target vulnerabilities that exist in systems due to the human factor. Many cyber attacks are spread via mechanisms that exploit weaknesses found in endusers, which makes users the weakest element in the security chain. The phishing problem is broad and no single silver-bullet solution exists to mitigate all the vulnerabilities effectively, thus multiple techniques are often implemented to mitigate specific attacks. This paper aims at surveying many of the recently proposed phishing mitigation techniques. A high-level overview of various categories of phishing mitigation techniques is also presented, such as: detection, offensive defense, correction, and prevention, which we belief is critical to present where the phishing detection techniques fit in the overall mitigation process.	Phishing has a list of negative effects on a business, including loss of money, loss of intellectual property, damage to reputation, and disruption of operational activities. These effects work together to cause loss of company value, sometimes with irreparable repercussions..

2.	WC-PAD: Web Crawling based Phishing Attack Detection	T Nathezhtha D Sangeetha V Vaidehi	October 2019	This paper proposes an automated customer care management system (CCM) to help maintain a good relation with customers. CRM can help any organization to survive and grow in a competitive market. It helps to know and treat each customer uniquely and effectively, resulting in a long-term fruitful relation with customer. This requires knowing the preferences of the individual customer. Making a successful CRM is very challenging as information about customer's preferences and behavior often difficult to obtain. In this paper we implemented a CRM system that can automatically communicate with present and future customers based on the information it has in its database. Making a database with the latest information about customer's trends and choice is crucial. This includes collecting data from various sources and then analyzing the data. Using modern computing techniques like data mining and web	It is easy to trick the crawler. Websites have hidden data that can be manipulated to make the page appear like it's something it's not.. Page rank can be manipulated. While search engine companies frown on the practice, there are ways to improve where your page appears on the list of results..

3 .	A Machine Learning Approach for URL Based Web Phishing Using Fuzzy Logic as Classifier	Happy Chapla Riddhi Kotak Mittal Joiser	July 2019	Phishing is the major problem of the internet era. In this era of internet the security of our data in web is gaining an increasing importance. Phishing is one of the most harmful ways to unknowingly access the credential information like username, password or account number from the users. Users are not aware of this type of attack and later they will also become a part of the phishing attacks. It may be the losses of financial found, personal information, reputation of brand name or trust of brand. So the detection of phishing site is necessary. In this paper we design a framework of phishing detection using URL.	The message contains malicious software targeting the user's computer or has links to direct victims to malicious websites in order to trick them into divulging personal and financial information, such as passwords, account IDs or credit card details.

4.	Spear-Phishing campaigns: Link Vulnerability leads to phishing attacks, Spear-Phishing electronic/UAV communication-scam targeted	Muhammed Sawood Baig Faisal Ahmed Ali Mobin Memon	November 2021	One of the most important strategies for gaining unauthentic early access to some person/company's computing resources/data is spear phishing. Phishing is, at its core, a sort of social engineering intended to persuade a user to give sensitive information or run a payload that will infect their system. Spear phishing is a type of phishing in which bogus emails are sent to specific businesses with the goal of obtaining confidential information. A successful phishing campaign necessitates the use of a few different resources as well as some setup. Impersonation, inducement, and access- control bypass techniques are among its approaches. In this paper we have studied and found up to date approaches to spear phishing attacks and their preventative measures. The paper also demonstrates the steps to set up and run successful phishing campaign and the results astonishingly shows that even only personality-targeted messaging can alter the response to phishing attacks. As a result of learning the facts, the paper suggests that users should seek to improve their security awareness by becoming familiar with the warning signs of phishing attacks. Moreover, more often in Unmanned Aerial Vehicles (UAV) or drones (which are now being used in various domains including military operations, monitoring, etc.), the resources are deployed as web services which makes them vulnerable to phishing activities. A data mining technique is also suggested as a tool for the detection of phishing attacks in UAVs.	The downside of this kind of attack is that, unlike regular phishing, this scam requires the hacker to spend a good amount of time on each victim. Each person much be researched in depth in order to gain trust, which is going to require more effort than simply sending out thousands of copies of the exact same scam email. What this means is that the end payout needs to be significantly higher than a regular phishing attack to make up for this extra time. This is why spear phishing is so much more dangerous, as a successful attack is going to steal even more money.

5.	HTMLPhish: Enabling Phishing Web Page Detection by Applying Deep Learning Techniques on HTML Analysis	Chidimma Opara Bo Wei Yingke Chen	July 2020	Recently, the development and implementation of phishing attacks require little technical skills and costs. This uprising has led to an ever-growing number of phishing attacks on the World Wide Web. Consequently, proactive techniques to fight phishing attacks have become extremely necessary. In this paper, we propose HTMLPhish, a deep learning based data-driven end-to-end automatic phishing web page classification approach.Specifically, HTMLPhish receives the content of the HTML document of a web page and employs Convolutional Neural Networks (CNNs) to learn the semantic dependencies in the textual contents of the HTML. The CNNs learn appropriate feature representations from the HTML document embeddings without extensive manual feature engineering. Furthermore, our proposed approach of the concatenation of the word and character embeddings allows our model to manage new features and ensure easy extrapolation to test data. We conduct comprehensive experiments on a dataset of more than 50,000 HTML documents that provides a distribution of phishing to benign web pages obtainable in the real-world that yields over 93% Accuracy and True Positive Rate. Also, HTMLPhish is a completely language-independent and client-side strategy which can, therefore, conduct web page phishing detection regardless of the textual language.	Phishers can also cost a company a significant part of its market value as a result of the loss of investors' confidence. Some investors would no longer trust the affected organization and may move their funds elsewhere to protect their portfolio.