Web Phishing Detection

Literature review

Introduction:

In recent times, Phishing becomes an important area of concern for security researchers because it is not difficult to develop the phishing website, which looks so close to legitimate website. Experts can identify phishing websites but not all the users can identify the phishing website and such users become the victim of phishing attack. Main aim of the attacker is to steal banks account details and personal information. In United States businesses, there is a loss of US$2billion per year because their clients become victim to phishing.

As per Index Report released in 2020, it was estimated that the annual worldwide impact of phishing could be as high as $1.6 million. Phishing attacks are becoming successful because lack of user awareness. Since phishing attack exploits the weaknesses found in users, it is very difficult to reduce them but it is very important to enhance phishing detection techniques.

In recent times, Phishing becomes an important area of concern for security researchers because it is not difficult to develop the phishing website, which looks so close to legitimate website. Experts can identify phishing websites but not all the users can identify the phishing website and such users become the victim of phishing attack. Main aim of the attacker is to steal banks account details and personal information. In United States businesses, there is a loss of US$2billion per year because their clients become victim to phishing. As per Index Report released in 2020, it was estimated that the annual worldwide impact of phishing could be as high as $1.6 million. Phishing attacks are becoming successful because lack of user awareness. Since phishing attack exploits the weaknesses found

in users, it is very difficult to reduce them but it is very important to enhance phishing detection techniques.

Literature review:

phishing detection and protection scheme (1):

Developing with the anti-phishing methods, phishers use various phishing methods and more complex and hard-to-detect approaches. The most straightforward way for a phisher to swindle people is to make the phishing web page similar to their target. However, many distinctive and features can distinguish the original legitimate website from the clone phishing website like the spelling error, image alteration, long URL address and abnormal DNS records. The full list is revealed in Table 3 which is used later in our analysis and classification study. If an attacker clones a legitimate website as a whole or designed to look similar as they usually do in most attacks in recent times,our approach is that similar looking phishing web page con-tent is not left for the users to check for the indicator or the authenticity attentively, but can detect by automated methods. Our approach is based on website phishing detection using the features of the site, content and their appear-ance. These properties are stored in a local database (Excel table) as a knowledge model and first compared with the newly loaded site at the time of loading against the dangerous web page offline. After the comparison was unable to detect the similarity, then the critical approach to compare the legitimate and fake using the features of the website with machine learning for an intelligent decision. The critical contribution of our approach includes Result:The output is determined by the classifier, in the phishing detection stage which predicts if the web page is suspicious, legiti-mate or phishing. The knowledge model and plug-in development will be developed at a later stage**.**

System detection related work (2):

Nowadays most people uses internet for various purposes such as online shopping like purchasing or selling products, chat with friends, sending mail. Internet users now spend more time on social networking sites Information can spread very fast and easily within the social media networks. Social media systems depend onusers for content contribution and sharing. Facebook had over 1.3 billion active users as of June 2014. there are over 1.3 billion (the number is keep growing) pages from various categories, such as company, product/service, musician/band, local business, politician, government, actor/director, artist, athlete, author, book, health, beauty, movie, cars, clothing, community. Fans not only can see information submitted by the page, but also can post comments, photos and videos to the page.

Result:

Domain anomaly features are used to identify possible malicious domains based on lexical and reputation factors, whereas social anomaly features represent anomalous user behaviors in social communications.

Learning to Detect Phishing Emails (3):

An alternative for detecting these attacks is a relevant process of reliability of machine on a trait intended for the reflection of the besieged deception of user by means of electronic communication. This approach can be used in the detection of phishing websites, or the text messages sent through emails that are used for trapping the victims. Approximately, 800 phishing mails and 7,000 non-phishing mails are traced till date and are detected accurately over 95% of them along with the categorization on the basis of 0.09% of the genuine emails.

Result:

We can just wrap up with the methods for identifying the deception, along with the progressing nature of attacks.

Phishing websites machine learning (4):

Phishing URL is a widely used and common technique for cybersecurity attacks. Phishing is a cybercrime that tries to trick the targeted users into exposing their private and sensitive information to the attacker. The motive of the attacker is to gain access to personal information such as usernames, login credentials, passwords, financial account details, social networking data, and personal addresses. These private credentials are then often used for malicious activities such as identity theft, notoriety, financial gain, reputation damage, and many more illegal activities. This paper aims to provide a comprehensive and comparative study of various existing free service systems and research-based systems used for phishing website detection. The systems in this survey range from different detection techniques and tools used by many researchers. The approach included in these researched papers ranges from Blacklist and Heuristic features to visual and content-based features. The studies presented here use advanced machine learning and deep learning algorithms to achieve better precision and higher accuracy while categorizing websites as phishing or benign. This article would provide a better understanding of the current trends and existing systems in the phishing detection domain.

Result:

Phishing URL detection plays a pivotal role for many cybersecurity software and applications. In this paper, we researched and reviewed works based on the advanced machine learning techniques and approaches that promise a fresh approach in this domain.

Support vector machine (5):

The existing anti-phishing approaches use the blacklist methods or features based machine learning techniques. Blacklist methods fail to detect new phishing attacks and produce high false positive rate. Moreover, existing machine

learning based methods extract features from the third party, search engine, etc. Therefore, they are complicated, slow in nature, and not fit for the real-time environment. To solve this problem, this paper presents a machine learning based novel anti-phishing approach that extracts the features from client side only. Below architecture diagram as shown in Fig. 1. represents mainly flow of training phase to Detection phase. First data need to be pre-processed and feature extraction using different feature sets and later we need to train this dataset with the corresponding algorithms and the output is displayed.

Result:

In future we can use a combination of any other two or more classifier to get maximum accuracy. We can also explore various phishing techniques that uses Lexical features.

REFERENCES

TITLE	PROBLEM IDENTIFIED	METHODOLOGY	STRENGTH	WEAKNESS
Phish Shield: A desktop application to detect phishing webpages through heuristic approach (Rao & ali,2015)	**To detect URL and website content of phishing pages**	**Heuristic approach**	**Ability to detect zero hour phishing attacks & increased speed in detecting speed in detecting phishing attack**	**High computational cost,inability to immediately update the whitelist & blacklist**
Mitigating cyber	**To tackle**	**Semantic content**	**Detecting of**	**It achieved 80%**

identity fraud advanced multi anti phishing technique(Yusuf et al,2013)	loopholes in electronic payment system security challenges in online banking transaction	analysis,Earth mover Distance(EMD) & biometric authentication with finger print	phishing webpages & preventing unauthorized online banking transfer & withdrawal	true negative
Efficient prediction of phishing website using supervised learning algorithms,Santhana Laksmi.v & vijaya MS,2011	Phishers are using new techniques to break all antiphishing mechanisms	Supervised learning algm,ie,.multi layer perceptron ,decision tree induction & machine learning techniques to model the prediction task & naive Bayes classification to explore result	It can predict whether a given website is legitimate or phishing website	Time taken to build the model & predictions accuracy is high in the case of decision tree induction
Anti phishing based on automated individual whitelist (AIWL)YE	Blacklist is not completely effective in detecting phishing URL because of	Naive bayesian classifier	It keeps a whitelist of users all familiar login user interface(LUIs)of website .it guides	It requires gathering the website IP & this is time consulting as IP needs to be

cao,Weli han & Yueran le,2008	**partial list of global phishing sites**		**against pharming attacks & low false positive**	**changed**