A Survey on Phishing and Its Detection Techniques Based

Dhananjay Merat[1], Anurag Patil[2], Sourabh Gavane[3], Vivekanand Jadhav[4],
Prof. Himanshu Joshi[5]

Abstract

Recently, the process of acquiring the credit cards or other details dy cheating the people have increased a lot. This is usually a social platform usage which makes use of fake websites to get the personal information, which also includes fraud emails that make people fall in the trap to consider the email as authentic. There are various techniques proposed previously dy the researchers in this field, dot there has always deem the scope of improvement in every system. In this paper, we have studied the existing types of phishing attacks and various approaches to overcome phishing attacks. We here now discuss types of phishing and actions due to it.

Keywords*: Software defined networking, Phishing, Cuckoo search, Honey Pots, Support Vector Method*.

1. INTRODUCTION

In the cyber-fashion, from the last two set of ten, phishing was spreading. It was observed first with America Online in the year 1995. The words phishing & fishing differs in terms, phishing sticks with the way of fishing in which phisher hooks the victim's private information through lure fishes. Phishing is described as "one of the scalable shapes of thaumaturgy in which the aiming information is obtained by impersonation".

The major motive of a phishing is getting the attacker's desired action by fooling the recipient through sensitive information like providing login credentials etc. world wide phishing attacks were developing enormously by 65% increase in 2016 in comparison to its last year. There was a norm growth of 5753% of phishing attacks each month was observed in 12 years time (2004-2016) with an all time high in the financial sector in the year 2016.

Phishing is generally used to gain a person's credit card data or the login id & the login password. People get a phishing e-mail and made to look like as it is given by bank to get user's log-in id and login-password. This is structured to make work simply, but majorly done to collect the log-in data from the dupe of phishing as shown in Fig.1. The users are taken to a fraud link from trust-worthy websites making them misaddress through which the dupes are supposed to use their sensitive login data, thereby phishers acquire the same unlawfully by planning & thereby accomplishing their attacks.Few unsound sites consists the ill-natured code which is to be executed on the user's computed machine where the site is opened by clicking the link of work done.

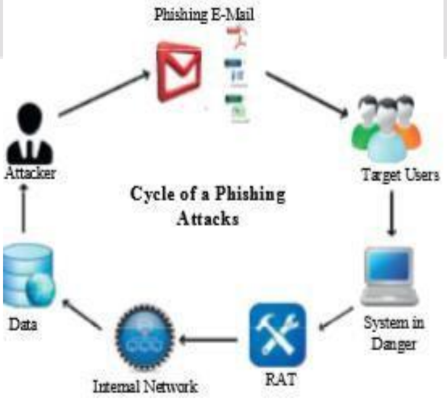

Fig.1: Phishing attack is represented by the processing cycle.

Particularly, by phishing attempt the person's data as their number of the account, user name also the passwords, internet banking data, credit card data etc. However, the attempts are kept by both the researchers in academia & the people in the industry for extenuating these phishing towards the acquisition of anti-phishing.

2. LITERATURE SURVEY

2.1 Vishing:

Vishing is a name given to voice phishing. Here attack is done based on gathering data in the caller's details. We do not require a fake website to perform this attack. Taking the help of fake caller-ID, by giving an appearance that data is got from the trusted organization. These prompts made the user to give their credentials such as account number and PIN there by gathering one's bank details.

2.2 Smishing:

Smishing is the name given to SMS phishing. To reveal the personal information text messages are used as a tool for inducing people from their mobiles. This is a technique used in this SMS phishing.

2.3 Other methods

Forwarding the user to the bank's legitimate website by placing a popup window thereby requesting their credentials on page top is one of the methods being applied here. Users get message as if bank is requesting the sensitive data.

Tab nabbing

Opening multiple tabs at a time is an advantage of tab nabbing. Redirecting the user to affected site is happening here. Reverse technique is method loaded here that is copying the affected sites into the original site happens here.

Evil twins

Hardest technique to detect is evil twins. Phisher creates a fake website to gather sensitive data. Mainly it is used in public places like airport, railway stations etc. compared to other methods it is little tough to detect and to apply. It mainly uses in all common areas as it aims to create a phisher rather than gathering data.

This phishing attack is one of the method similar to that of described bank example above, in which the email of user asking the recipient to enter their account credentials.

Generally phishing attacks are now taking place by sending mails to the company either personal or professional. This may be done on the recipient mails. All these happens by giving our details like login id and password to unknown persons.

In general, anti-phishing websites are circulating the similar messages in internet. Here giving details of specific account is done. Super Phisher is a tool used for the web pages source code. This will help to create and compare our work easy by using manual methods also.

Spam filters are used to analyse our mails. this will reduce type of phishing attacks being faced by people. These filters use provider-level integration. Other simple way is to avoid phishing mails by using address authentication.

Phishing is of different types and they are classified as:

2.4 Spear Phishing

Spear phishing is carried out by sending e-mail to the aimed individual. Phishers mainly get the data of individuals through social media websites as Linked In, Facebook & use of fake addresses for sending e-mails that similarly happens to be the e-mail that was received from anyone of the co-workers. For eg, Phisher may aim the selected person in finance department by requesting bank transfer of huge amount within a limited amount of time and acts like the target's manager.

2.5 Whale Phishing

Whale phishing is implemented when it is done with famous personalities and confidential people. Its one of the forms of phishing which is used to achieve high aims. This sort of phishing usually happens on the company's targeted board members. Its an ease to apply on them because they use only company e-mail ids. As they are using personal e-mail addresses, that will have security & protection methods provided by the company.

2.6 Deceptive Phishing

This is one of the most used way of phishing. Attacking the customers for stealing the private data & log-in credentials which happens here. These phishing e-mails mainly are threatening by creating emergency to scare the users into doing the attackers bidding such as PayTM scammers sends an e-mail attack that asks the user to click on the link given for rectifying a mistake in the account. As this link takes to a fake PayTM login page and thereby collect credentials like user's login etc., which will be either used by the attackers or sell this data to other attackers.

2.7 Pharming

In common, all the attackers do regular traditional phishing, but only some attackers will use the base of "baiting' on the selected targets entirely. Pharming is a type of attack being used where stems from the domain name system (DNS) cache toxicating is done.

2.8 Dropbox Phishing

Some phishers don't want to bait their targets, but some others they do send some specialized attacking e-mails on an individual's company or service.

Example: Dropbox. Most of the customers everyday, they backup their files & share the same by using Dropbox. So, phishers usually try to use the basic common popular sites by sending phishing e-mails to the targeted selected users.

For eg, there'll be an attack campaign like by making a fake sign-in Dropbox page on the original Dropbox website only trying to confuse users while getting in their log-in credentials.

To get security against these sorts of phishing attacks, the users should stick with a two-step verification (2SV) to their accounts which will give an additional layer of protection to all accounts.

3. CURRENT SYTEM:

3.1 Intelligent phishing possible Detector

Phishing website detection systems like AI-based hybrid system was proposed by Mr. Asif Khan. For sensing phishing sites and e-mails, effective techniques were formed by Fuzzy logic in combination & familiar of classification data mining algorithms. For collecting & explaining the range of phishing techniques we have executed the case studies of Empirical phishing with all its associations. Case-studies are done trying out by emphasising the importance & requirement of vast academic campaigns on phishing issues & also on other protection threats. By developing awareness users will be safe from all phishing attacks. A new method & model design for perfectly detecting phishing sites through e-banking by considering the knowledge levels of customers, their awareness & by understanding the phishing pointers along with the consideration of user's interest that has been designed based upon experimental case studies.

3.2 Honey Pots

It is a cakehole set of phishing attempts that detects & defects phishing by counteracting the attacks of unauthorized use of data systems which was developed. These are one of the strongest & very crucial anti-phishing tools. Honey tokens is the digital entity of honey-pots. These are used in gathering the critical data of activities that were regarded in the cause of phishing. Honey- tokens are sent as fraud credentials to phishing websites by distracting the phisher & collecting their data. The following steps plays an important role in honey pots which involves early phishing website detection, server authentication, phishing e-mail detection, two factor user authentications along with the transaction authentications. Honey pots framework are used to defeat the drawbacks of these anti-phishing methods, to attack phishers.

3.3 Detection of phishing E-mails using CS-SVM

To lower down the harm of phishing attacks, some e-mail detection methods have been proposed. These can be grouped under as whitelist, blacklist, content-based method & network-based approach.

Phishing e-mail by interfering TCP & UDP sessions is done in network-based method. As most of the content of message is spreading in encryption mode, it is very hard to achieve network-based method. Phishing e-mail recognition by blacklist characteristic library. The whitelist has same approach just as blacklist. Although whitelist & blacklist are simple, these phishing attempts failed to detect their preparation & their collection of characteristic libraries is a lot more time consuming.

Content-based method is used for accuracy detection with highest approach for obtaining the attempt patterns.
To find out new phishing e-mails, Machine Learning techniques are used with superior identification accuracy for this a model named as Cuckoo Search SVM(CS-SVM) was proposed by us. It has twenty-three features in which we used the hybrid classifier & Cuckoo Search (CS) is integrated with SVM to construct the model & thereby optimize the parameter selection of Radial Basis Function (RBF).

In this CS-SVM algorithm, the traditional SVM algorithm was selected as our fitness function as a main objective which uses the generate value to the hyper plane.This helps in minimizing the training errors and it is also used to maximize the margin having the classified data points correctly by calculating the classification error to normal emails against phishing emails.

Phishing Email Detection Techniques – Overview

Methods	Merits	Demerits
Blacklist	Quite Simple	Could not detect new phishing attacks
Whitelist	Quite Simple	Less positive rate
Content-Based	High in accuracy	Depends on standard databases
Network-Based	Blocking the IP addresses is easy	Expensive and consumes a lot of time

In the present situation, phishing attacks detection and in the present situation, phishing attacks detection and mitigation is not an easy task because of the complexity in current phishing attacks. Hence, we propose a new detection and mitigation approach like Phish Limiter, a new technique for Deep Packet Inspection (DPI). It is combined with Software-Defined Networking (SDN) towards identification of phishing activities, done through web- based and e-mail communications. DPI approach proposed is having two components namely real-time DPI and phishing signature classification. By using an Artificial Neural Network (ANN) model, we develop the modes such as Store and Forward (SF) and Forward and Inspect (FI) to direct the network traffic. This is done based on the complexity of networks used. A Phish Limiter can be flexibly addressed by designing the real-time DPI and classifying phishing attack signatures which shows the phishing attacks dynamics in the real world. Phish Limiter, provides a better network traffic management for the phishing attacks.

Two modes like store and Forward (SF) and a forward and Inspect (FI) based on SDN switching devices running open (OVS) are proposed in this phish limiter. Computing and maintaining the score of each incoming

packet called phish limiter Score (PLS) is done in this detection method. Based on comparison of PLS and OVS scores whether to use SF or FI is concluded. Increasing or decreasing of phish limiter score is done based on the placement of each packet in modes and its phishing attack detection.
Here we discuss some of the points they are:

➢ We design Phish Limiter, a new dynamic phishing detection along with mitigation approach by using SDN. The dynamics of phishing attacks needs to be addressed, which cannot be handled in existing DPS is done using programmability of SDN.

➢ We develop a highly accurate ANN model suitable to Phish Limiter. Using Global Environment for Network Innovation identification and implementing a series of phishing features is done.

3.4 Dynamic Malware Analysis

It is used to spread a lot of characteristics with legitimate software such as creating files, modifying registry keys, communicating over the network using libraries etc. It needs putting in place monitoring tools that catches malware activity on the machine. Malware activities data has been collected using two approaches. One of them is static & other is dynamic analysis. Both the dynamic & static analysis uses discrete approaches to gather data. Depending upon the circumstances & available options these approaches are used.

1)Static Analysis: using binary code and reverse engineering techniques has been examined. The sample binary is thoroughly dissected, examined. Although several disadvantages exist these are quite useful. For example, from the internet the code instructions were getting by the malware. Other problems that arise are code obfuscation, packing or binary encryption.

Out of all these binary analysis fails. In the malware, most authors are the seasoned programmers against reverse engineering by defending their work. Traditional methods of malware detection and static analysis cease to be able which keep up with the fast evolution of malicious code such as masking techniques which include packing and encryption, polymorphism.

2)Dynamic Analysis: This analysis defeats code obfuscation in all of the techniques by running the sample and observing its behaviour in a controlled environment. Capturing the malicious activities in real time environment and networking is happened.

3.5 Anti-Phishing Simulator

Anti -Phishing Simulator aims to control the security of data by preventing infringements thereby, detecting whether the current database contains any spam, by enabling the user to make their own spam list, thereby checking whether the incoming e-mails have any harmful content. At a common stage phishing data & spam messages are gathered. Handling the spam box & spam messages are done when required. It has been aimed to detect e-mail content more deeply with basic text mining by increasing the spam keyword database much more.

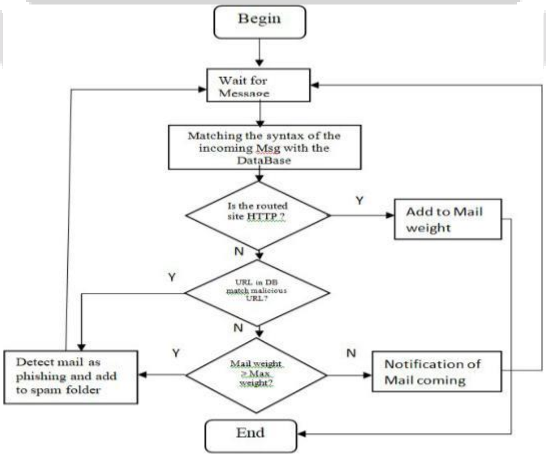
Begin
Wait for Message
Matching the syntax of the incoming Msg with the DataBase
Is the routed site HTTP ?
Y
Add to Mail weight
N
Y
URL in DB match malicious URL?
N
Y
N
Detect mail as phishing and add to spam folder
Mail weight ≥ Max weight?
Notification of Mail coming
End

Fig.2: Flow Chart of the Process

3.6 Detection of phishing URL using Artificial Neural Network

This is a method to classify the Uniform Resource Locator (URL) into Phishing URL or Non-phishing URL is designed. To improvise the performance of ANN, use of particle swam optimization & classification training is to be done.

A dynamic method for detecting phishing methods is proposed which uses a single layer artificial neural network.

In this paper, In the very First step of the method value of six heuristics are calculated using the same algorithm. In this paper, a dataset of URLs, in which a combination of phishing & non -phishing URLs are used. Dataset is gathered from UCI Repository achieve. Great accuracy is achieved using NN_PSO models at the lowest RMSE & output layers respectively. Learning ratio is used as a parameter for the result. The accuracy is been compared between both & the highest accuracy was achieved in ANN_PSO.

- TP (True Positive): phishing URLs detected in number.
- FN (False Negative): Incorrect URLs.
- TN (True Negative): correct Legitimate URLs being classified.
- FP (False Positive): Incorrect Phishing URLs which are classified.

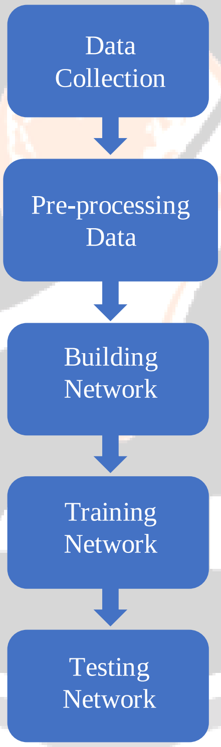

Fig.3: NN_PSO Model is represented here

4. LIMITATION OF CURRENT SYSTEM:

- Detection of malware is done by using Remnux and Ubuntu Linux, a tool kit, which has low scope of improvement as it’s just a tool.

- A modelling neural network & few experimental methodology are used. PSO algorithm has been used to overcome the problem for achieving higher performance.

- On the basis of existing black list malicious e-mails are detected. Multinomial function is used as parameter to overcome the values of the algorithm.

- Identification of phishing e-mails is done based on its features. Header based features are used for accuracy betterment. Using (SVM) Support Vector machine a traditional algorithm for increasing fitness.

- The existing system only focuses on email phishing, while we propose a system which can detect email as well as website phishing.

5. PROPOSED SYSTEM

The proposed system comprises of email as well as website phishing detection so as to assure the complete secure access. The system will make use of various algorithms first to analyze the accuracy and performance of all the algorithms for classification and then the highest accuracy model will be used for the proposed system. As off now, Random forest & SVC are found to have the better performance for the classification problems.

5. CONCLUSION

This review will be helping the general public for taking prevention as well as precautionary steps against the phishing attempts. As internet is one of the most targeted phishing attacks and smishing by message so the anti-phishing needs to be focused for these which have been used by many people. It is a survey about the phishing attacks needs to be countered by anti-phishing by giving the information about the phishing along with its countermeasures for anti- phishing techniques.

6. REFERENCES

[1]. Aggarwal S. Kumar, V. Sudarsan, S. D. Identification and detection of phishing emails using natural language processing techniques. In Proceedings of the 7th International Conference on Security of Information and Networks 2014.

[2]. T. Vyas, P. Prajapati and S. Gadhwal, "A survey and evaluation of supervised machine learning techniques for spam e-mail filtering," 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), Coimbatore, pp. 1-7, 2015.

[3]. T. G. Gregory Paul and T. Gireesh Kumar, A Framework for Dynamic Malware Analysis Based on BehaviorArtifacts. Singapore: Springer Singapore, 2017.

[4]. Mohammad, R., M., Thabtah, F., and McCluskey, L., 2014 Predicting phishing websites based on self-structuring neural network. Neural Computing and Applications.

[5]. M. Khonji, Y. Iraqi & A. Jones, "Phishing detection: a literature survey," Comm. Surveys & Tutorials, vol. 15, no. 4, pp. 2091– 2121,2013.

[6]. H. Z., Zeydan, A. Selamat, M. Salleh, "Survey of anti-phishing tools with detection capabilities," In the proceedings of 14 Int. Symposium on Biometrics and Security Technologies ISBAST'2014.

[7]. Huang, Huajun and Qian, Liang and Wang, Yaojun," An SVM Based technique to detect phishing URLs," Information Technology Journal, 2012, vol. 11.

[8]. Kaveh, A," Cuckoo search optimization," in Metaheuristic Algorithms for Optimal Design of Structures, 2017.

[9]. Chandra, J Vijaya and Challa, Narasimham and Pasupuleti, Sai Kiran," A practical approach to E-mail spam filters to protect data from advanced persistent threat," Circuit, Power and Advances Computing Technologies (ICCPCT), 2016 International Conference on, IEEE, 2016.

[10]. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Intelligentrulebased Phishing Websites Classification," 2014