# A Survey on Phishing And It's Detection Techniques Based on Support Vector Method (SVM) and Software Defined Networking(SDN)

**A. MahaLakshmi, N. Swapna Goud, Dr. G. Vishnu Murthy**

***Abstract: Phishing is a deceitful attempt for obtaining the sensitive information like credit card details, user names and passwords. It is one of the social engineering methods that gathers personal information through websites such as malicious websites and deceptive e-mail to canvass personal information from a company or an individual by prance as a trustworthy entity or organization. Phishing often attacks email by using as a vehicle and even sending messages by email to users that represent a part of a company or an institution who perform business such as financial institution, banking etc. Phishing is becoming more malicious day by day and its detection is very important. In cyberspace, phishing is motivating the researchers to develop the model through which we can develop more security towards the safe services provided by the web. Here we discuss types of phishing and conflicts due to it.***

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

## I. INTRODUCTION

In cyber world, from the past two decades, phishing was plaguing. It has been reported first with America Online in 1995. Both the words phishing and fishing varies in terms but phishing follows the way of fishing in which phisher lures the victim's personal information through bait and fishes. Phishing is defined as "one of the scalable forms of deception in which the target information is obtained by impersonation".

The main aim of a phishing is taking the attacker's desired action by tricking the recipient through sensitive information like providing login credentials etc. Globally phishing attacks were growing tremendously by 65% increase in 2016 comparing to its previous year. There was an average increase of 5753% of phishing attacks per month was reported in 12 years time (2004-2016) with an all time high in the financial sector in 2016.

Phishing is used often to know the persons credit card information or login id/password. User receives a phishing e-mail as if it is given by bank to take user's login id and password. This is designed to make work easy, but mainly done to collect the login information from the victims of phishing as shown in Fig.1. The users are taken to a spoofed link for trustworthy websites making them mislead through which the victims are supposed to use their credentials thereby phishers acquire the same illegally by planning and thereby executing their attacks. Some of the fraudulent websites contains the malicious code which needs to be executed on the user's machine where website is opened by clicking the link of work done.

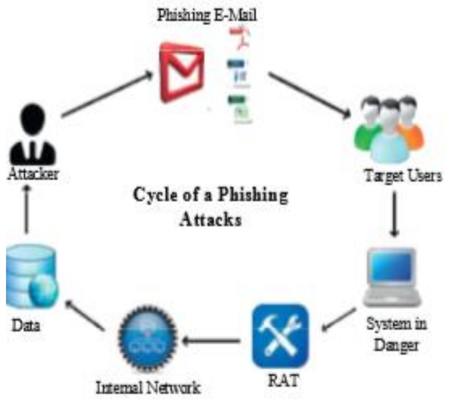

**Fig.1: processing cycle of phishing attack.**

In general, by phishing attack the user information such as their account number, user name as well as passwords, internet banking information, credit card information etc. However, the efforts are being kept by both researchers in academia and the industry people for mitigating these phishing towards achievement of anti-phishing.

## 2. TYPES OF PHISHING ATTACKS

### 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 organisation. These prompts made the user to give their credentials such as account number and PIN there by gathering one's bank details.

### 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.

**Revised Manuscript Received on December 22, 2018.**

**Mahalakshmi,** Post Graduate Student, Department of CSE, Anurag Group of Institutions, Hyderabad, T.S. India.(Email: mahalakshmi0525@gmail.com)

**N Swapna Goud,** Assistant Professor, Department of CSE, Anurag Group of Institutions, Hyderabad, T.S. India.(Email: swapnagoudcse@cvsr.ac.in)

**Dr.G. Vishnumurthy,** Professor, Department of CSE, Anurag Group of Institutions, Hyderabad, T.S., India.(Email: hodcse@cvsr.ac.in)

### 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 method 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:***

### Spear Phishing

Spear phishing is done by sending mail to a targeted individual. Phishers generally got the information of individuals through social media sites such as Linkedin, Facebook and use of fake addresses for sending emails that similarly happens to be the mail that was received from anyone of our co-workers. For example, Phisher may target the selected person in finance dept. by requesting bank transfer of large amount within a short time and acts like a victim's manager.

### Whale Phishing

Whale phishing is applied when it is done on big personalities and confidential people. It is one of the forms of phishing used to achieve high targets. This type of phishing generally happens on company targeted board members. It is very easy to apply on them as they use only company mail id. As he is using personal email address, that will have security and protection features giving by company.

### Deceptive Phishing

This is one of the most common ways of phishing. Attacking the customers for stealing the personal information and login credentials happens here. These phishing mails generally threaten by creating urgency to scare the users into doing the attackers bidding such as PayTM scammers, sends an email attack that asks the user to click on the link provided for rectifying a mistake in their 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.

### Pharming

In General all attackers do normal traditional phishing, but only some attackers will use the idea of "baiting' on the selected victims entirely. Pharming, a type of attack being used where stems from domain name system (DNS) cache poisoning is done.

### Dropbox Phishing

Some phishers don't want to bait their victims, but some others they do send specialized attack emails on an individual company or service.

**Example:** Dropbox. Millions of customers everyday they backup their files and share the same by using Dropbox. So, phishers generally try to use the common popular sites by sending phishing mails to the target selected users.

For example, there will be an attack campaign like by creating a fake sign- in Dropbox page on the original Dropbox site only trying to confuse users while entering their login credentials.

To get protection from all these type of phishing attacks, the users should follow a two-step verification (2SV) to their accounts which will provide an additional layer of security to all our accounts.

## 3. PHISHING DETECTION TECHNIQUES:

### Intelligent phishing possible Detector

Mr. Asif Khan had proposed phishing website detection systems like AI-based hybrid system. For detecting phishing websites and mails, efficient techniques were formed by Fuzzy logic in combination and association of classification data mining algorithms. For gathering and explaining the range of phishing methods we have implemented the case studies of Empirical phishing with all its relations. Case-studies are done experimentally by emphasising the importance and requirement of vast educational campaigns on phishing problems and also on other security threats. By creating awareness customers will be safe from all phishing activities. A new approach and model design for perfectly detecting phishing websites through e-banking by considering the knowledge levels of users, their awareness and by understanding the phishing pointers along with consideration of user's interest that has been designed based on experimental case studies.

### Honey Pots

It is a trap set of phishing attacks that detects and defects phishing by counteracting the attempts of unauthorized use of information systems which was created. These are one of the powerful and very important anti-phishing tools. Honey tokens is digital entity of honeypots. These are used in collecting the critical information of activities that were involved in the cause of phishing. Honeytokens are sent as fake credentials to phishing sites by confusing the phisher and collecting their information. The following steps plays a crucial role in honey pots which includes early phishing site detection, server authentication, phishing mail detection, two factor user authentication along with transaction authentications. Honey pots framework are used to overcome the drawbacks of these anti-phishing techniques, to attack phishers.

### Detection of phishing E-mails using CS-SVM

To reduce damage of phishing attacks, some email detection techniques have been proposed. These can be grouped under as whitelist, blacklist, content-based approach and network-based approach.

Phishing email by interfering TCP and UDP sessions is done in network based approach. As most of the content of message is transmitting in encryption mode, it is very difficult to implement network-based approach. Phishing email recognition by blacklist characteristic library. The whitelist has same method like blacklist. Although whitelist and blacklist are simple, these phishing attacks failed to detect their preparation and their collection of characteristic libraries is very time – consuming.

Content-based approach is used for accuracy detection with highest approach for obtaining the attack patterns.

To trace out new phishing emails, Machine Learning techniques are used with high identification accuracy For this a model named Cuckoo Search SVM(CS-SVM) was proposed by us. It has 23 features in which we used the hybrid classifier and Cuckoo Search (CS) is integrated with SVM to construct this model and 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

| Techniques | Advantages | Disadvantages |
|---|---|---|
| **Blacklist** | simple | Cannot detect new phishing attacks. |
| **White list** | simple | Positive rate is less. |
| **Content-based** | Highest accuracy. | Depends on standard databases. |
| **Network-based** | Easy to block IP addresses. | Time consuming and costly. |

### Phish Limiter using Software- Defined –Networking

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.

### Dynamic Malware Analysis

Malware is used to share a lot of characteristics with legitimate software like creating files, modifies registry keys, communicates over the network, uses libraries etc. This requires putting in place monitoring tools that captures malware activity on the machine.

Malware activities information has been gathered using two approaches.one of them is static and other is dynamic analysis. Both dynamic and static analysis uses different approaches to collect data. Depending on the circumstances and available options these methods 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 was 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

### Anti-Phishing Simulator

Anti -Phishing Simulator aimed to control the security of information by preventing infringements thereby checking whether the current database is having any spam, by enabling the user to create his own spam list, thereby checking whether the incoming mails has any dangerous content. At a common point phishing data and spam messages are collected. Controlling spam box and spam messages when required is done. It has been aimed to analyse mail content more thoroughly with basic text mining by increasing the spam keyword database much more.

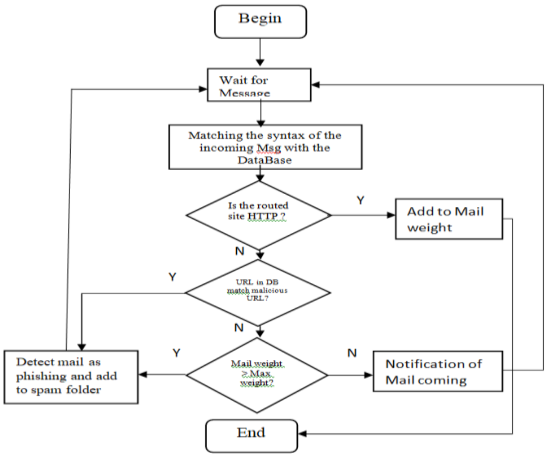
Begin
Wait for Message
Matching the syntax of the incoming Msg with the DataBase
Is the routed site HTTP ?
Y
Add to Mail weight
N
Y
URL in DB match malicious URL?
N
Y
Detect mail as phishing and add to spam folder
Mail weight ≥ Max weight?
N
Notification of Mail coming
End

**Fig.2: Flow Chart of the Process**

### Detection of phishing URL using Artificial Neural Network

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

A dynamic approach for detecting phishing techniques is proposed which uses a single layer artificial neural network. In this paper,

In the First step of the technique value of six heuristics are calculated using this algorithm..

In this paper, a dataset of URLs, in which a combination of phishing and non -phishing URLs are used. Dataset is collected from UCI Repository achieve. Good accuracy is achieved using NN_PSO models at the lowest RMSE and output layers respectively. Learning ratio is used as a parameter for the result. The accuracy has been compared between both and 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.

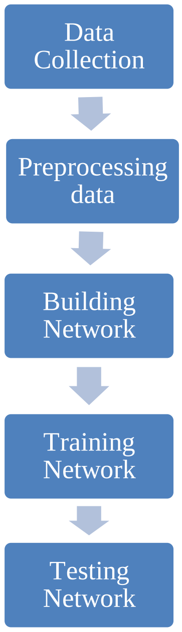

**Fig.3: Represent NN_PSO Model**

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

## 4. RESULTS

| Protocol | Highlighting Features | Requirements | Weakness or over head |
|---|---|---|---|
| Mohammad Abu Qbeitah∗, and MontherAldwairi | Analysis of phishing e-mails is done dynamically.<br>Malicious samples behaviour is investigated in controlled environment. | Analysing samples is done by setting up two laboratories.<br>Honey net is used for real samples.<br>To analyse dynamic samples in real environment. | Detection of malware is done by using Remnux and Ubuntu Linux, a tool kit. |
| Surbhi Gupta,AbhishekSinghal | A method to classify Uniform Resource Locator(URL) into phishing URL or non-phishing URL is given.<br>Particle swam optimisation has been used for training ANN. | An unrecognised website with incorrect URL has been used for proper authentication.<br>Tab napping, a technique used by attacker to disclose the personal data.<br>One of the techniques is e-mail. | A modelling neural network and some experimental methodologies are used.<br>PSO algorithm has been used to overcome the problem for achieving higher performance. |
| TianruiPeng,Ian G. Harris,YukiSawa | To analyse text has been done to evolve phishing attacks using natural processing.<br>Large set of e-mails are used to increase effectiveness. | A detection algorithm has been given to processes each sentence at a time and returns true if it has an attack. | On the basis of existing black list malicious emails are identified. Multinomial function is used as parameter to overcome the algorithm values. |
| MuhammetBaykara,ZahitZiya Gürel | The attacker sends malicious e-mails in the form of messages which has been termed as cyber-attacks. Anti-phishing simulator is given for software purposes. | For the purpose of evaluation basis training defence attacks is a parameter.<br>URL of address bar in java script is processed. | Punctuality process has been focused in the process of an improved threat attack.<br>Using various spam filtering techniques and various features presented a small set of values are found. |
| WeinaNiu, Xiaosong Zhang, Guowu Yang, Zhiyuan Ma, ZhongliuZhuo | Support Vector Machine (SVM) is more effective in the study of machine learning concept. | Cuckoo Search SVM(CS-SVM) is designed to improve accuracy.<br>CS-SVM extracts 23 features, based on it a hybrid classifier is modelled.<br>Cuckoo Search (CS) has been integrated with SVM for optimizing Radial Basis Function(RBF) | Detection of phishing e-mails is done based on its features.<br>Header based features are used for accuracy improvisation.<br>Using supportvector machine (SVM) a traditional algorithm for increasing fitness. |

## 5. CONCLUSION

This review will help the general public for taking prevention as well as precautionary steps against the phishing attacks. As internet is one of the most targeted phishing attack and smishing by message so the antiphishing 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.

## 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,"A SVM Based technique to detect phishing URLs," Information Technology Journal, 2012, vol. 11.
8. Kaveh, A,"Cuckoo search optimization," Advances 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 Computing Technologies (ICCPCT), 2016 International Conference on, IEEE, 2016.
10. R. M. Mohammad, F. Thabtah, and L. McCluskey, "Intelligentrulebased Phishing Websites Classification," 2014