Different Types of Phishing Attacks and Detection Techniques: A Review

Kiran D. Tandale
Dept of Electronics & Telecommunication's Engg.
MGM's Jawaharlal Neharu Engineering College
Aurangabad, India
kirantandale592@gmail.com

Dr. Sunil N Pawar
Associate professor, Dept of Electronics & Telecommunication's Engg
MGM's Jawaharlal Neharu Engineering College
Aurangabad, India
sunilpawar@jnec.ac.in

***Abstract*— In this paper a review has been taken on different types of phishing attacks and detection techniques. Most of the cyber attacks are spreading through users weaknesses, which makes user weakest elements in security chain. Mostly this phishing targets the vulnerability which is already in the system due to human factor. Phishing is the huge problem and there is no single solution for mitigating the vulnerability completely so one has to use number of techniques to mitigate the attack. Here we presented various categories of mitigation technique for the phishing website like detection, correction, prevention and offensive defense.**

Keywords— ***Phishing Attacks, Web Sites, Authentication, Security***

I. INTRODUCTION

Phishing attacks is the combination of social engineering and the technology for gaining the restricted access of information of end user. Now a day's one of the common method is used for phishing is mass emailing to victim from website of some authority. These websites are little bit similar with the online bank, electronics payment sites, virtual marketplace like flipkart and government agencies. This fraud website collects the information from the user in the name of security, account update or authentication. Sometimes they download malware on clients machine instead of collecting information. Phishing is specially designed for stealing important personal data like debit, credit card information, credentials or any other personal information. In a single day millions of such emails are sent to the victims which are similar to the site which you trust and asking for credit, debit card details.

Recently phishing detection has led to the growth of numerous new visual similarity based approaches. Visual similarity based approaches compare the visual appearance of the suspicious website to its corresponding legitimate website by using various parameters. Due to different phases of phishing detection, this paper contains the following:

1. Anatomy of phishing attack presents phases of attacks and types of phishing attacks.

2. Related works represents the phishing attack history, financial losses in world. This section describes overall phishing picture of phishing attacks.

3. Next, in phishing attack and it's impact section, we describe how the phishing attacks happened in world and how the phishing word derived. In this section we also mention where did financial losses due to phishing attacks and what was the impacts of phishing on economy and reputation of firms.

4. In next section it is about the how we can detect the phishing attacks.

5. In anti-phishing method section it states the different methods are given which can help us from phishing attacks. There we mention four different method to protect from phishing attack.

6. After that it compare the seven different algorithms of detection of phishing. In comparison we compare contribution summary, weakness, mechanism and algorithm.

7. And finally it concluded about phishing attacks and detection techniques.

II. ANATOMY OF PHISHING ATTACKS

There are three phases of phishing attack: first one is, get a phish to a victims, next is victim gets alert to take action which is given in message, most of the time it is regarding to visit the website or install malware; and the third one is illegal selling of stolen data. Most of the time this type of phishing emails use social technique rather than the technical to fool the users. First need of people is recognized by phished and then they misguide the people. example is, an warning message appeared on screen to a system administrator regarding to upgrade the security patch and then instead of patch an malware is being installed.

There are total eleven types of phishing attacks as search engine phishing, man-in-man-middle phishing, content-Injection, DNS-based phishing, Data theft, host file poisoning, Deceptive phishing, malware-based phishing, key logger & screen capturing, session hacking and web Trojans.

III. RELATED WORK

Chernobyl virus was the first virus who attack on embedded software [1]. The main objective of that malware was to erase all data of hard disk and reset the BIOS at some other date. Mobile phone was also become the target for worms [2] & [3] found in home routers, wireless access points and switches. Some threads was found in open firmware by Adelstet Kozen and Stillerman. So to minimize the probability boot software are in stack based language.

Hence, code analyzer check malicious code at loading and abstract aged code from running. Smith, Farber and Arbaugh implement a cryptographic access control system, AEGIS where only bootstrapping firmware installation is allowed. This paper will explore the various phishing, where the phishing is spread through online market place and the hardware spoofed by malicious embedded software. The domestic network router which is attacked by malware steals

978-1-7281-5970-6/20/$31.00 ©2020 IEEE

the information by passive eavesdropping and DNS spoofing.

IV. Phishing Attack and It's Impacts

Phishing is social engineering type of attack where criminals spreads SMS and emails among the peoples and get their information or install malware on PC. These messages are so much similar to the popular brands and electronics market places [4]. Phishing works very cleverly, then it doesn't matters either it is big organization or an individuals security. Phishing doesn't matters encryption software, firewall, two fact authentication mechanism of organization or individuals have.

"Phishing" this word originate from the analogy which cyber criminals used emails lures to "phish" for the financial data and passwords from the ocean of internet users. In terminology use of "Ph" is partly lost in history, but some of famous hackers called it as "Phreaks" which was hackers used in hacking of telephones(Phreaking). Then in 1996 this tern coined by the hackers who stole America Online(AOL) users password data. Internet of phishing become popular when it was mention inalt.2600 hacker news group in January [5].

After some time, this definition of phishing attack expanded and not remain limited up to account details, but now a days it includes the access of financial as well as personal data. Before phishing was done by replying to emails for credits card details and password but now it expanded and includes fake websites, screen capturing, man-in-the-middle data proxies, installation of trojen horses key loggers.

Now a days phishing is classified into two categories, exploit based phishing and web based phishing. Attacks exploits well known vulnerabilities in famous web browsers for installation of malwares into users computers, this comes under exploits based phishing. Usually illusive phishing can be classified into coping images in URL or using illusive text [4]. Victims may becomes fooled by making fake websites with very minor differences in spell of domain name, for example difference between "l" and "i" hence "paypal.com" and "paypai.com" may seems similar. Phishers also use non-printing characters and no-ASCII Unicode characters [4].

V. Detection of Phishing Attacks

Now a days, we use Wi-Fi for internet connectivity in home and office place. This type of network is very easy to use and connects different types of devices. But this wireless connection may damage network security and user information. Although secure wireless connection may reduce the risk while connecting to device to wireless network, many wireless networks are still unsecured [4]. Now, phishers used wireless connection to trick user to get their information. Evil twin attack like treats may found in wireless network, so for that it is advice to clones a clients preset Set Service Identifier (SSID) to identify theft or any other malicious activity.

There are some methods which are used for detection of phishing attacks,

A. Statistical Based Method

This method is characterised by having explicit underlying probability model, which is able to provide the probability that an website belongs in each class rather than classification. In statistical based method there are two algorithm, one is Naive Bayesian network(NB) and other is Bayesian Network.

1) Bayesian Network

It is the model where graphical representation is used for the probability relationship among set of variables. The Bayesian network structure S is a directed acyclic graph (DAG) and the nodes in S are in one-to-one correspondence with the features X. The arcs represent casual influences among the features while the lack of possible arc in S encodes conditional independencies. Moreover, a feature (node) is conditionally independent from its non descendants given its parents. Typically, the task of learning a Bayesian network can be divided into two subtasks: initially, the learning of the DAG structure of the network, and then the determination of its parameters. Probabilistic parameters are encoded into a set of tables, one for each variable, in the form of local conditional distributions of a variable given its parents.

2) Naive Bayesian network

Naive Bayesian networks (NB) are very simple Bayesian networks which are composed of DAGs with only one parent (representing the unobserved node) and several children (corresponding to observed nodes) with a strong assumption of independence among child nodes in the context of their parent. The major advantage of the naive Bayes classifier is its short computational time for training. In addition, since the model has the form of a product, it can be converted into a sum through the use of logarithms with significant consequent computational advantages.

B. Logic-Based Methods

In this section we will concentrate on of logical (symbolic) learning methods: decision trees, random forest and c5.0.

1) Decision Tree

Decision tree induction is the learning of decision trees from class-labeled training tuples. Adecision tree is a flowchart-like tree structure, where each internal node (non leaf node)denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node. The construction of decision tree classifiers does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery. Decision trees can handle high dimensional data. Their representation of acquired knowledge in tree form is intuitive and generally easy to assimilate by humans. The learning and classification steps of decision tree induction are simple and fast. In general, decision tree classifiers have good accuracy.

2) Random Forest (RF)

Random forest (RF) is an ensemble learning classification and regression method suitable for handling problems involving grouping of data into classes. The algorithm was developed by Breiman and Cutler. In RF, prediction is achieved using decision trees. During the training phase, a number of decision trees are constructed (as defined by the programmer) which are then used for the class

prediction; this is achieved by considering the voted classes of all the individual trees and the class with the highest vote is considered to be the output.

C. Kernal Method

Kernel Methods are best known for the popular method Support Vector Machines which is really a constellation of methods in and of it. Kernel Methods are concerned with mapping input data into a higher dimensional vector space where some classification or regression problems are easier to model.

1) Support Vector Machines (SVM):

In formal definition, a support vector machine design a hyperplane or set of hyperplanes in a high or infinite dimensional space, which can be used for classification, regression or other tasks. A SVM is a promising new method for classification of both linear and nonlinear data. Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. Support vector machine algorithms divide the n dimensional space representation of the data into two regions using a hyperplane. This hyperplane always maximizes the margin between the two regions or classes. The margin is defined by the longest distance between the examples of the two classes and is computed based on the distance between the closest instances of both classes to the margin, which are called supporting vectors.

D. ANN-Based Method

Artificial Neural Networks are models that are inspired by the structure and/or function of biological neural networks. They are a class of pattern matching that are commonly used for regression and classification problems but are really an enormous subfield comprised of hundreds of algorithms and variations for all manner of problem types.

1) Neural Networks:

An artificial neural network, or neural network, is a mathematical model inspired by biological neural networks. In most cases it is an adaptive system that changes its structure during learning. There are many different types of NNs. For the purpose of phishing detection, which is basically a classification problem, we choose multi layer feedforward NN. In a feedforward NN, the connections between neurons do not form a directed cycle. Contrastedwith recurrent NNs, which are often used for pattern recognition, feedforward NNs are better at modeling relationships between inputs and outputs.

VI. Anti-Phishing Methods

There are four methods Anti- phishing attack which are given below,

A. Use a Custom DNS Service

You have to use DNS resolution service so the you can easily access the website which you wants. A machine don't know the wed address, so that it ask to DNS resolution service for the IP address of the same web address. There are independent DNS companies where do much than name resolution. That companies can filters the sites based on contents and malware/phishing concern [6].

B. Use Browser's Phishing list

Browser check the website to visit with the phishing websites list, then if it found as phishing then browser gives pop-up on screen[6,7].

C. Use Sites to Check link

There are some online links where you can check about the website includes malware and phishing or not [8].

D. Use your own Skill

When you got the emails like "you have won prize" and "bumper offer" then you first consider it is a phishing.

VII. Comparative View of Existing phishing detection Algorithms

Author	Contribution Summary	Weaknesses	Mechanisms	Algorithms
Chandra Sekaran, et. al. [9]	Structural Feature	-Small size datasetb -200 email only -time consuming	prototype implementation site between (MUA) and (MTA)	Support vector machine (SVM) Classifier
Ganger et. al. [10]	Training smart screen	-Recall measurement level is low -Works on fix features	Uses feedback databases of Microsoft	Bayesian statistics 100,000 emails attributes
Bazargani gilani [11]	Phishing email's classification by heuristic method, concept of Semantic ontology	-Accuracy is low as compare to other technique	Model work in 5 steps	concept of Semantic ontology for th TFV method Information gain(IG), Nave Bayes algorithm classifiers
Chandra Sekaran Chinchani et.al. [12]	PHONEY: Mimicking user response	-Collected data is very small in size -Time consumption is more	PHONEY: Technique is installed Between MUA and MTA	PHONEY: Mimicking user response
Feishtte, Sadeh et. al. [13]	PILFERS Prototypes	-Big Size number of phishing emails not classified properly	WHOIS query has ten different features	support vectors machines (SVM) and Randon forest classifiers used
Bergholz et. al. [14]	Study the statically filtering of phishing emails	-More features, -Time consumption is more -Memory Requirement is high	Train the classifier by features obtain based on dynamic Markov's chain and Class- Topic Model	Dynamic Markov chain and a class Topic Models
Ma otoghi et. al. [15]	Robust classifier model	-Few features are used -uses Non-standard Dataset	7 hybrid features, Model consist of 5 stages appears	Information gain algorithms, Decision tree algorithm, C4.5

CONCLUSION

Day by day phishers are finding the different obstacle for phishing detection and developing their technique to manipulates communication and try to increase the phishing case. Already there are many precautions to avoid phishing such block listing, although this may happen. There are so many precautions and preventive measures which organizations implements to ensure their network and websites remain secure. This review approaches the many anti phishing methods among these methods machine learning approaches are considered as most effective method to detect phishing. This method ensure 100% accuracy.

REFERENCES

[1] CERT. Incident note IN-99-03.http://www.cert.org/incident notes/IN-99-03.html, April 1999.

[2] Ivan Arce. The shellcode generation. IEEE Security & Privacy, September/October 2004.

[3] Ivan Arce. The rise of the gadgets. IEEE Security & Privacy, September/October 2003.

[4] Stajano, F. and Wilson, P. Understanding scam victims: Seven principles for systems security. Commun. ACM 54, 3 (Mar. 2011), 70–75.

[5] Hong, J. Why have there been so many security breaches recently? Blog@CACM.

[6] Purkait, S., 2012. Phishing counter measures and their effectiveness – literature review, Information Management & Computer Security, 20 (5), pp.382 – 420.

[7] Jagatic, T.N., Johnson, N.A., Jakobsson, M., and Menczer, F. Social phishing. Commun. ACM 50, 10 (Oct. 2007), 94–100.

[8] Anderson, R., 2001. Security Engineering: A Guide to Building Dependable Distributed Systems, London, Wiley Computer Publishing.

[9] Chandrasekaran, et. al., "Phishing email detection based on structural properties", in New York state Cyber Security Conference (NYS)," 2006.

[10] P.R. D.L Ganger, "Gone phishing evaluation anti phishing tools for windows, technical report, "September 2006.

[11] M.Bazargranigilani, "Phishing detection using Ontology concept and Nave Bayes A0lgorithm, "International general of Research and reviews in computer science, Vol. 2, no.2, 2011.

[12] M. Chandrasekaran, et. al., " Phoney, Mimicking user response to detect phishing attacks," in In: Symposium on World of Wireless, Mobile and Multimedia Networks, IEEE Computer Society, 2006 PP. 668-672.

[13] I. Fette, et al., " Learning to detect phishing Emails," in proc. 16th International World Wide conference (WWW 2007) ACM press, New York, NY, USA, May 2007, pp 649-656.

[14] A. Bergholz, et al, "Improved phishing detection using model based features," in proc, conference on Email and Anti-Spam (CEAS) Mountain view conf, aug 2008.

[15] I, Ma, et al, "Detecting Phishing emails using hybrid features,"IEEE Conf 2009," pp. 493-497".

[16] Tyler Moore, "Cooperative attack and defense in distributed networks" Technical Report, UCAM-CL-TR-718, ISSN 1476-2986.