International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 4 Issue 01,January-2015

A Survey on various Phishing and Anti-Phishing Measures

Ms. Neha R. Israni
Dept of Computer Science & Engineering
GHRIETW
Nagpur, Maharashtra

Mr. Anil N. Jaiswal
Dept of Computer Science & Engineering
GHRIETW
Nagpur, Maharashtra

***Abstract*— Phishing, a semantic attack which targets the user rather than the computer is turning into a breeding ground for vast fraudulency over the internet; and therefore is one of the most challenges toward internet security. Phishing is a way in which perpetrators adopt social engineering schemes by sending e-mails, instant messages or online advertising to allure users to phishing websites that mimic trustworthy websites in order to trick individuals into revealing their sensitive information like, financial account details or passwords which can then be exploited for a specific reason. To protect users against phishing attacks various anti-phishing strategies have been implemented having different strategies.**
This paper presents the various phishing attacks followed by a review of various anti-phishing techniques.

***Keywords*— *Phishing, Anti-phishing, Internet Security, URL*.**

I. INTRODUCTION

Phishing is a widespread attack which targets the user rather than the computer [1]. The term phishing was first described in detail in 1987, and the first registered use of the term "phishing" was made in 1996. Phishing is a way of trying to acquire sensitive information such as usernames, passwords and credit card details, by pretending as a trustworthy entity in an electronic communication. This is similar to Fishing, where the fisherman puts a bait on the clip, thus, pretending to be a genuine food for fish. But the clip inside it takes the complete fish out of the lake. Data exchange arising from popular social web sites, online payment processors or IT administrators is commonly used to lure the unsuspecting public.

Phishing activity is generally carried out by either e-mail spoofing or instant messaging and it often directs users to enter details at a fake website whose look and feel are almost identical to the legitimate one [6]. Phishing is an example of social engineering techniques used to deceive users, and exploits the poor usability of current web security technologies.

Figure 1 shows the phishing process, where a phisher allures the user to enter sensitive information on a fake website.

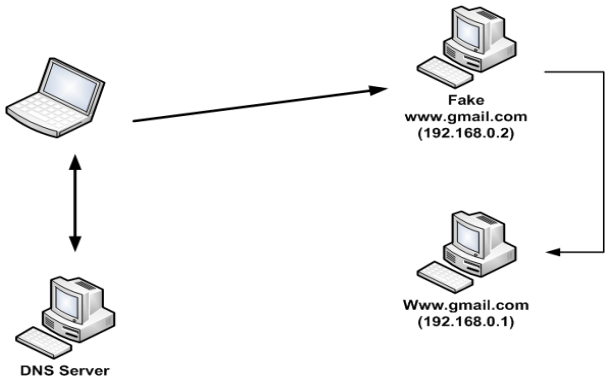

Fig 1. Phishing Process

According to a phishing activity trend report published by Anti-phishing working group in the first quarter of 2014, the number of phishing sites hoped by 10.7 percent over the fourth quarter of 2013 [2]. Payment Services continued to be the most-targeted industry sector at the beginning of 2014, with 46.51 percent of attacks during the three-month period [2], which can be seen in figure 2.

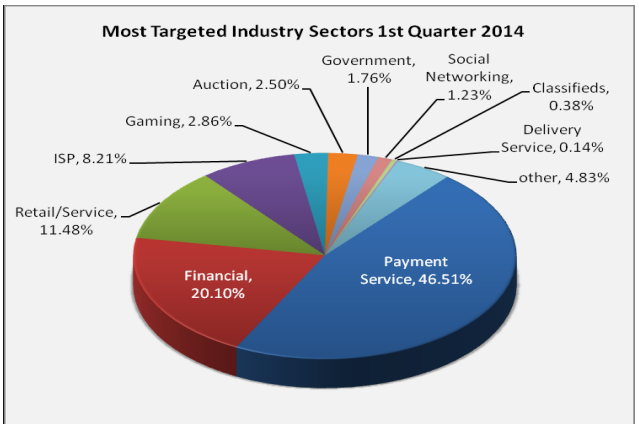

Fig 2. Industry Sector wise effect of Phishing [2]

Various attempts to deal with the growing number of reported phishing incidents include legislation, user training, public awareness, and technical security measures which can be carried out either at sever or at client side.

This paper presents a survey of the research related to the various phishing and anti-phishing methods. The rest of the paper is organized as follows: Section II reviews research work related to A) Phishing Strategies, B) Anti-Phishing Measures. Section III covers the Literature Survey of some Heuristic Anti-Phishing Methods. Section IV concludes the paper.

II. RELATED WORK

A. *Phishing Strategies*

The main technologies of the phishing site follow the following procedure: l) How to distribute the information of the phishing sites to users; 2) How to trick users to make them think this is a legitimate Web site. Most phishing sites use two ways to trick the users, 'Similar Spoofing' [3] and 'Instant Spam Messages' [3].

1) Similar Spoofing [3]: This technique is used by the phishers, where the pages of the phishing sites are very similar to the legal pages. That is, the URL of the phishing sites is identical to the URL of legitimate sites and so the users will be misguided towards conceiving a phishing site as a legitimate Web site.

2) Instant Spam Messages [3]: Phishers send spam messages to the users notifying them that there are some problems in their internet banking and asks them to reenter the password to solve these problems. The phishers send links which are displayed in the mail which look like the URL of the internet banking, but indeed they are anchored to a phishing site.

However, a third unrecognizable attack 'DNS-Based Phishing' [4] practiced by phishers is to tamper with industry's domain name system so that requests for URLs or name service return a fake address and ongoing communications are directed to a fake site.

B. *Anti-Phishing Strategies*

Over the past years, a great deal of work has been done on anti-phishing resulting in various anti-phishing methods. Some techniques work on emails, some aid on attributes of websites while some on the URL of the websites.

The main anti-phishing techniques can be classified as follows:

1) List Based Methods [5]: These methods use an ever-updating list of phishing websites. Most commercial tools like browsers and security toolbars use this approach.

Advantage - The benefit of this method is that it requires low computational cost. Also it results in 100% accuracy on decision about websites that are present in the blacklist and produce less false positives.

Disadvantage- The main drawback is that it returns 0% accuracy when not in the blacklist. Another drawback can be large memory overhead.

2) User-Based Methods [5]: These methods involve user in the decision making by informing the users and lets the user decide the action.

Advantage- A knowledgeable user who is aware about recent Phishing attack can bifurcate in correct path.

Disadvantage- Requires skilled people for decision making. Also the decision can result in large false positives

3) Heuristic Based Methods [10]: Heuristic anti-phishing detection methods apply some heuristic rules to decide whether a site is phishing site or not. The heuristic approach may involve:

a) Analyzing URL
b) Analyzing Text Contents
c) Analyzing Visual Similarity
d) Analyzing Image Insertion
e) Analyzing content from the redirected Web page

Most of the heuristics are subjective and involve different mechanisms [10]. They can evolve from URL matching till the actual page layout. A few Heuristic approaches are discussed below.

III. LITERATURE SURVEY

In spite of a lot of work that has been done on implementing better and efficient tools for phishing detection and prevention, still it is very difficult to completely eradicate the problem and to find the actual number of users that are caught as victim. A few methods are explained and distinguished here.

A. *Ensemble Clustering for Internet Security Applications [1]:*

This method first finds the associated Web Pages with the given page, then mines its features (such as link relationship, ranking relationship, webpage text similarity, and web page layout similarity relationship) between the given web page and its associated WebPages [1], and then applies DBSCAN clustering algorithm to decide if there is a cluster around the given webpage. If such cluster is found, the given webpage is regarded as a phishing webpage; otherwise, it is identified as a legitimate webpage.

Advantage- Studies have shown that it performs well and detects a large number of phishing sites against various anti-phishing toolbars.

Disadvantage- For metamorphic phishing websites, some website level constraints should be used.

B. *Hybrid System to Find & Fight Phishing Attacks Actively [5]:*

It is a hybrid method that detects general phishing attacks in an active way through DNS query logs and known phishing URLs [5]. A phishing repository is used which is composed of confirmed phishing URLs, to get the most possible phishing paths. And to it are applied various heuristic rules to improve the detection.

Advantage- The method is effective and efficient in detecting general phishing activities. According to APAC's feedback, almost all detected reports by this system are confirmed as phishing attacks resulting in a detection accuracy around100%.

Disadvantage- The only drawback of this system is that the DNS query logs can record all the websites only if they are visited once by DNS server users.

C. An Entice Resistant Automatic Phishing Detection [6]:

This system utilizes supervised offline machine learning, where, in the learning phase, a collection of websites are provided that are marked as either legitimate or phishing [6]. These websites are then given to the feature extractor. And then the whole collection of vectors is fed to ML engine which generates a classifier.

Advantage- The system produce much better results in terms of accuracy against Cantina and Cantina+ against which it was compared.

Disadvantage- The system is yet to be improved to reduce the false positive rate.

D. Data Shield Algorithm (DSA) for Security against Phishing Attacks [7]:

The Data Shield Algorithm is 'Real-time Light weighted Anti-Phishing Algorithm' that can detect and prevent the users from phishing attacks in the real time [7]. The algorithm detects fraudulent web site using Proper DNS and IP matching by performing all the functionalities automatically.

Advantage- The algorithm was proved efficient enough to detect and prevent both the known and unknown phishing attacks. It is also capable of Domain name protection and Webpage appearance protection.

Disadvantage- Again under this algorithm, for metamorphic phishing websites, some website level constraints should be used.

E. Intelligent Phishing Website Detection and Prevention System by Using Link Guard Algorithm [8]:

Link Guard is light weighted and can detect and prevent phishing attacks in real time. Link Guard works by analyzing the differences between the visual link and the actual link. It also calculates the similarities of a URL with a known trusted site [8].

Advantage- LinkGuard is effective to detect and prevent both known and unknown phishing attacks with minimal false negatives.

Disadvantage- LinkGuard may result in false positives, since using dotted decimal IP addresses instead of domain names may be desirable in some special circumstances. In some cases LinkGuard may result in false negatives. False negatives are more harmful than false positives.

F. Phish Mail Guard [10]:

In the proposed system a hybrid method is used for phishing mail detection, which is a combination of blacklist, white list and heuristic technique [10]. For heuristic technique, textual and URL analysis are done for further classification. The Phishing mail detection algorithm works by analyzing the blacklist and whitelist checking, difference between the visual link and the actual link, textual analysis and lexical analysis.

Advantage- The effectiveness of detection is increased as lexical URL analysis is used. Also the false positive rates are less.

Disadvantage- The time consumed for detection is more for unknown phishing attacks.

IV. CONCLUSION

An extensive variety of researches have been made on protecting users from the phishing attacks. Each work has its own technique, contribution and limitations. In this paper, we attempted to provide a comprehensive survey on various heuristic methods that have been implemented and are being practiced. The "Cluster Ensemble Technique" helps in classifying whether a site is phishing or not by cluster formation method, a similar approach of machine learning is practiced in "An Entice Resistant Automatic Phishing Detection". The methods like "A Hybrid System to Find & Fight Phishing Attacks Actively" and "Data Shield Algorithm" work on DNS matching. While the "Link Guard Algorithm" and "Phish Mail Guard" works on the URL analysis. As a survey paper, we may not include each and every aspect of individual works; however a wide variety of approaches have been focused to create a vivid understanding of work done in anti-phishing area. An effort has also been made to design and implement a "web request interceptor to prevent phishing attack" that works on network socket programming and the DNS address resolution protocol in order to get the details about URL and the IP addresses provided by user while internet surfing. This is a multiple level authentication scheme for phishing detection.

V. REFERENCES

[1] Weiwei Zhuang, Yanfang Ye, Yong Chen, and Tao Li, "Ensemble Clustering for Internet Security Applications" in IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS NOVEMBER 2012

[2] Phishing Activity Trends Report 1st Quarter 2014, Published June 23, 2014 \ www. apwg.org

[3] Yang Liu, Miao Zhang,"Financial Websites Oriented Heuristic Anti-Phishing Research' In *Proceedings of Ieee Ccis2012*

[4] Gaurav, Madhuresh Mishra,Anurag Jain, "Anti-Phishing Techniques: A Review", In International Journal of Engineering Research and Applications" ISSN: 2248-9622 Mar-Apr 2012, pp.350-355.

[5] Hong Bo, Wang Wei, Wang Liming, Geng Guanggang, Xiao Yali, Li Xiaodong, Mao Wei, "A Hybrid System to Find&Fight Phishing Attacks Actively" in 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology

[6] Hossain Kordestani, Mehdi Shajari, "An Entice Resistant Automatic Phishing Detection" in 2013 5th Conference on Information and Knowledge Technology (IKT)

[7] Ram Avtar1, Bhumica Verma2 and Ajay Jangra3, "Data Shield Algorithm (DSA) for Security against Phishing Attacks" in Research Cell: An International Journal of Engineering Sciences ISSN: 2229-6913 Issue Sept 2011, Vol. 4.

[8] M.MADHURI1, K.YESESWINI2, U. VIDYA SAGAR3, "Intelligent Phishing Website Detection And Prevention System By Using Link Guard Algorithm" in International Journal of Communication Network Security, ISSN: 2231 – 1882, Volume-2, Issue-2, 2013

[9] Weiwei Zhuang, Qingshan Jiang, Tengke Xiong, "An Intelligent Anti-phishing Strategy Model for Phishing Website Detection" in 2012 32nd International Conference on Distributed Computing Systems Workshops.

[10] Jayshree Hajgude, Lata Ragha, "Phish Mail Guard :Phishing Mail Detection Technique by using Textual and URL Analysis" in *2012 World Congress on Information and Communication Technologies*