V.S.B.ENGINEERING COLLEGE, KARUR

Department of Electronics and Communication Engineering

IBM NALAIYA THIRAN

LITERATURE SURVEY

TITLE : **Web Phishing Detection**

DOMAIN NAME : **Security**

LEADER NAME : Abulhasan A

TEAM MEMBER NAME: Asikahamed Y

Ajithkumar B

Gokulram D

MENTOR NAME : Janani S

ABSTRACT

Web service is one of the key communications software services for the Internet. Web phishing is one of many security threats to web services on the Internet. Web phishing aims to steal private information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate entity. It will lead to information disclosure and property damage. This paper mainly focuses on applying a deep learning framework to detect phishing websites. This paper first designs two types of features for web phishing: original features and interaction features. A detection model based on Deep Belief Networks (DBN) is then presented. The test using real IP flows from ISP (Internet Service Provider) shows that the detecting model based on DBN can achieve an approximately 90% true positive rate and 0.6% false positive rate.

INTRODUCTION

Web service is a communication protocol and software between two electronic devices over the Internet [1]. Web services extends the World Wide web infrastructure to provide the methods for an electronic device to connect to other electronic devices [2]. Web services are built on top of open communication protocols such as TCP/IP, HTTP, Java, HTML, and XML. Web service is one of the greatest inventions of mankind so far, and it is also the most profound manifestation of computer influence on human beings [3].

With the rapid development of the Internet and the increasing popularity of electronic payment in web service, Internet fraud and web security have gradually been the main concern of the public [4]. Web Phishing is a way of such fraud, which uses social engineering technique through short messages, emails, and WeChat [5] to induce users to visit fake websites to get sensitive information like their private account, token for payment, credit card information, and so on.

The first phishing attack on AOL (America Online) can be traced back to early 1995 [6]. A phisher successfully obtained AOL users personal information. It may lead to not only the abuse of credit card information, but also an attack on the online payment system entirely feasible.

The phishing activity in early 2016 was the highest ever recorded since it began monitoring in 2004. The total number of phishing attacks in 2016 was 1,220,523. This was a 65 percent increase over 2015. In the fourth quarter of 2004, there were 1,609 phishing attacks per month. In the fourth quarter of 2016, there was an average of 92,564 phishing attacks per month, an increase of 5,753% over 12 years [7]. According to the 3rd Microsoft Computing Safer Index Report released in February 2014, the annual worldwide impact of phishing could be as high as $5 billion [8].

LITERATURE SURVEY

Research background and related works

Phishing attacks are categorized according to Phisher's mechanism for trapping alleged users. Several forms of these attacks are keyloggers, DNS toxicity, Etc., [2]. The initiation processes in social engineering include online blogs, short message services (SMS), social media platforms that use web 2.0 services, such as Facebook and Twitter, file-sharing services for peers, Voice over IP (VoIP) systems where the attackers use caller spoofing IDs [3, 4]. Each form of phishing has a little difference in how the process is carried out in order to defraud the unsuspecting consumer. E-mail phishing attacks occur when an attacker sends an e-mail with a link to potential users to direct them to phishing websites.

Conclusion

The proposed study emphasized the phishing technique in the context of classification, where phishing website is considered to involve automatic categorization of websites into a predetermined set of class values based on several features and the class variable. The ML based phishing techniques depend on website functionalities to gather information that can help classify websites for detecting phishing sites. The problem of phishing cannot be eradicated, nonetheless can be reduced by combating it in two ways, improving targeted anti-phishing procedures and techniques and informing the public on how fraudulent phishing websites can be detected and identified. To combat the ever evolving and complexity of phishing attacks and tactics, ML anti-phishing techniques are essential. Authors employed LSTM technique to identify malicious and legitimate websites. A crawler was developed that crawled 7900 URLs from AlexaRank portal and also employed Phishtank dataset to measure the efficiency of the proposed URL detector. The outcome of this study reveals that the proposed method presents superior results rather than the existing deep learning methods. A total of 7900 malicious URLS were detected using the proposed URL detector. It has achieved better accuracy and F1—score with limited amount of time. The future direction of this study is to develop an unsupervised deep learning method to generate insight from a URL. In addition, the study can be extended in order to generate an outcome for a larger network and protect the privacy of an individual.

REFERENCES

1. https://en.wikipedia.org/wiki/Web_service.
2. O. Adam, Y. C. Lee, and A. Y. Zomaya, "Stochastic resource provisioning for containerized multi-tier web services in clouds," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, no. 7, pp. 2060–2073, 2017.

3. T. Bujlow, V. Carela-Espanol, J. Sole-Pareta, and P. Barlet-Ros, "A survey on web tracking: Mechanisms, implications, and defenses," *Proceedings of the IEEE*, vol. 105, no. 8, pp. 1476–1510, 2017.

4. H.-C. Huang, Z.-K. Zhang, H.-W. Cheng, and S. W. Shieh, "Web application security: Threats, countermeasures, and pitfalls," *The Computer Journal*, vol. 50, no. 6, pp. 81–85, 2017.