ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

Detection of Phishing Website Using Machine Learning

Ritika Sakpal[1], Shubham Shirke[2], Aditya Shinde[3], Tejas Shenoy[4]

Diploma students, Computer Department, Thakur Polytechnic, Maharashtra, India

ABSTRACT: Phishing is a new type of network attack where the attacker creates a replica of an existing web page to buffoon users in to submitting personal, financial, or password data to what they think is their service provider's website and believe it is authorized. The concept is an end-host based anti-phishing project for avoiding phishing attacks. The project is the concept for finding the phishing website to grasp the information of the end user. Project is based on the careful analysis of the phishing website. Each end-user can detect the phishing website and can be safe guarded from a phishing attack. After doing so the end-user recognizes the phishing website and can avoid browsing such harmful sites. The user can check if the website is malicious from the database. The project uses the ASP.net as front-end and C# and SQL database management as back-end.

KEYWORDS: Phishing website detection, System design, Machine learning.

I. Introduction

Phishing is a new term produced from the word 'fishing', it refers to the act that the attacker lure users to visit a fake website by developing a look alike website, and stealthily get users personal information such as username, password, financial details, account details, national security ID, etc.

This information obtained than can be used for future target advertisements or even identity thefts, attacks (e.g., transfer of money from one's account). The frequently used attack method is to send e-mails, messages, which can cause theft of data or personal information.

Passwords of social networking accounts, credit cards are miss-entered every day, or the attackers are providing upgrading services, to lure you to visit their website to conform and modify your personal information through the fake websites.

If you input the data i.e. your personal information, the attackers than successfully collect it on the server side, and is able to perform their next step actions with that obtained information of yours and use it for their malicious intentions. Phishing itself is a new concept, but it's increasingly used by the attackers i.e. the phishers to steal your personal information and perform business and social crimes in recent years. Within four to five years the number of phishing attacks have increased dramatically. Phishing attacks are commonly used and are easy to execute on their target.

II. LITERATURE SURVEY

Phishing website detection techniques are broadly classified into two categories, user education and software. In user education approach user has to be educated about the safe browsing practices. Software approach has different machine learning based techniques. Some of these techniques are explained in detail here

III. Proposed System

There are number of websites who ask users to provide sensitive information such as username, password or credit card details etc. often for malicious reasons. 6,35,000 e-banking websites have been attacked by phishing websites which is an enormous number which cannot be neglected.

These type of websites are known as phishing websites. In order to detect and predict phishing websites, an intelligent, flexible and effective system that is based on using Machine Learning algorithm is proposed. Such phishing websites can be detected based on some important characteristics like URL Domain Identity and detecting blacklisted

DOI: 10.15680/IJIRCCE.2018. 0601020

keywords in the final phishing detection rate. With the help of this website users can prevent getting scammed by such malicious websites. This website can be used by everyone in order to prevent themselves from being a victim by the hands of the attackers. Machine Learning algorithm used in this system will help the users to achieve better performance and efficiency as compared to other classification algorithms. With the help of this system users can also purchase products online and enter information without any hesitation.

IV. HOW IT WORKS?

- This system uses effective machine learning algorithm to detect the phishing websites.
- The phishing websites can be detected based on some important characteristics like URL and Domain Identity and blacklisted keywords.
- All the secure website's data will be entered into the database which will be surfed when the user enters any URL into the check website box.
- It will also check the blacklisted keywords present into the URL Domain Identity.
- The user has the ability to give suggestions for some unnoticed, blacklisted websites which will be monitored by the administrator and if found malicious it would be added to the blacklist.
- There will be a separate panel dedicated for news updates related to phishing attacks and also "Did you know?" stuff which will enlighten the user about what measures could be taken on their level to stay secure from phishing website attacks.
- Administrator has the control over the website which will ensure the websites uploaded in the blacklist that could lead to a phishing attack. The administrator also ensures that no websites which are genuine and safe for surfing the internet are affected by adding to the blacklist.

V. FLOW DIAGRAM

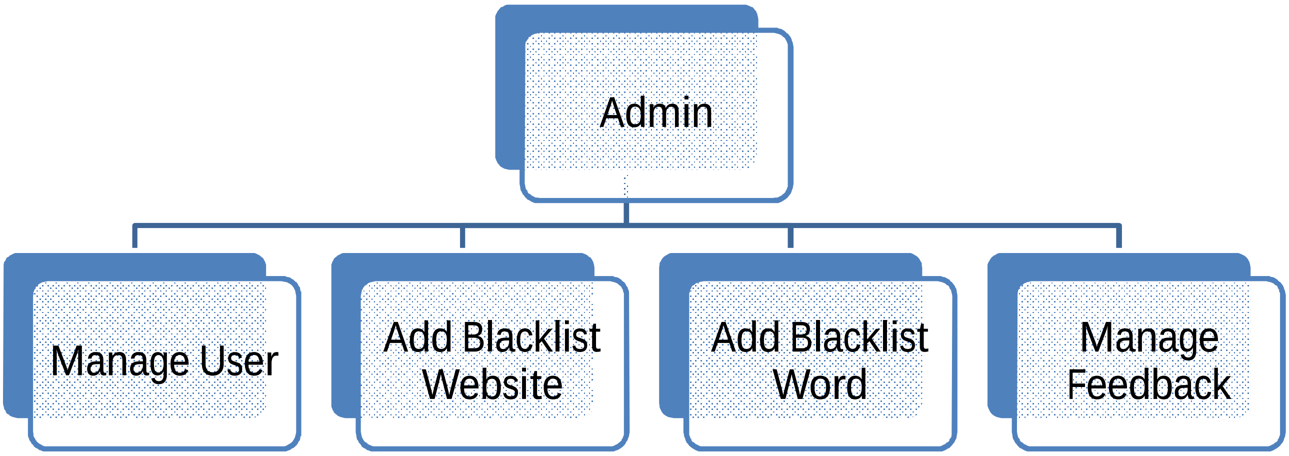

Fig. 1. Admin Panel

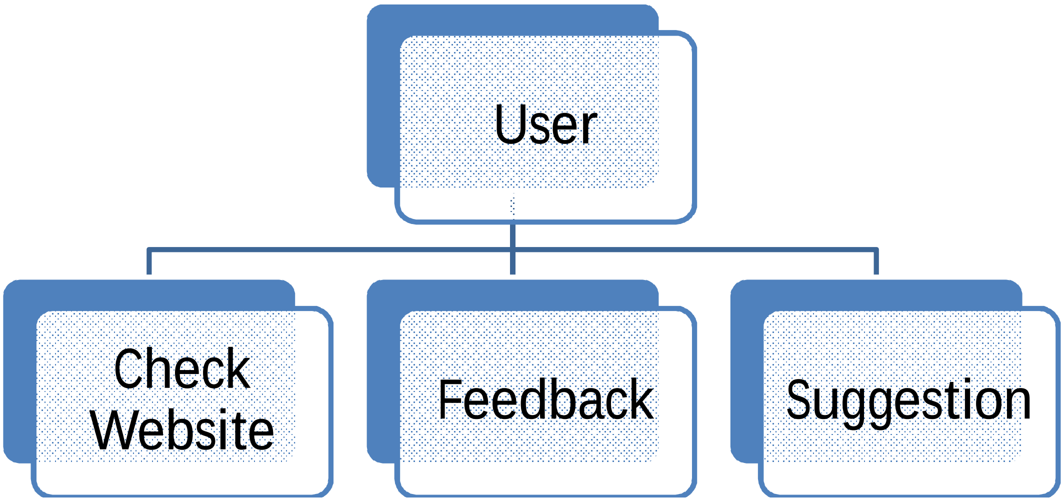

Fig. 2.User Panel

VI. ADVANTAGES

- This system can be used by everyone in order to achieve safe surfing over the internet.
- Payments can be done securely over the internet on secure and genuine e-banking sites.
- This helps to build good customer relationship which would benefit the customers as well as the E-commerce websites.
- Machine Learning algorithm used in this system provides better performance as compared to other traditional algorithms.
- With the help of this system users can also purchase things and sets online without any hesitation.

  User can also give suggestion for new phishing sites, this sites will be monitored by the admin and if found malicious it would be added in the database.

VII. FEATURES

- Easy to maintain.
- Efficient.
- Reliable.
- Low Cost.
- User Friendly.
- Feedback system.

VIII. CONCLUSION

It has become a serious network security problem, facing financial loss of billions of dollars to both consumers and the e-commerce companies. And perhaps more eventually, phishing has made e-commerce distrusted and attractive to normal consumers. In this paper, we have modified the characteristics of the hyperlinks that were modified in phishing e-mails and messages.

We then designed an anti-phishing algorithm, upon the derived characteristics.

The motive of this website made by us is to provide secure surfing on internet so that in the upcoming future the users feel safe and secure to access the free universe of internet. In the coming years, India is looking forward to become a "Digital India" and our project could play an important role in building a "Secure Digital India".

REFERENCES

1. CALMAN, C. Bigger phish to fry: California’s antiphishing statute and its potential imposition of secondary liability on internet service providers. Richmond Journal of Law and Technology XIII, 1 (2006).
2. CAVUSOGLU, H., AND RAGHUNATHAN, S. Configuration of Detection Software: A Comparison of Decision and Game Theory Approaches. Decision Analysis 1, 3 (2004), 131–148.
3. CRANOR, L. F. A framework for reasoning about the human in the loop. In UPSEC’08: Proceedings of the 1st Conference on Usability, Psychology, and Security (Berkeley, CA, USA, 2008), USENIX Association, pp. 1–15.
4. DHAMIJA, R., AND TYGAR, J. D. The battle against phishing: Dynamic Security Skins. In SOUPS ’05: Proceedings of the 2005 symposium on Usable privacy and security (New York, NY, USA, 2005), ACM Press, pp. 77–88.
5. DOWNS, J., AND FISCHHOFF, B. Adolescent Health: Understanding and Preventing Risk Behaviors. John Wiley and Sons, 2009, ch. 5.
6. DOWNS, J. S., HOLBROOK, M. B., AND CRANOR, L. F. Decision strategies and susceptibility to phishing. In SOUPS ’06: Proceedings of the second symposium on Usable privacy and security (New York, NY, USA, 2006), ACM Press, pp. 79–90.
7. EGELMAN, S., CRANOR, L. F., AND HONG, J. You’ve been warned: an empirical study of the effectiveness of web browser phishing warnings. In CHI ’08: Proceeding of the twentysixth annual SIGCHI conference on Human factors in computing systems (New York, NY, USA, 2008), ACM, pp. 1065–1074.
8. FLYNN, J., SLOVIC, P., AND MERTZ, C. K. Gender, race, and perception of environmental health risks. Risk Analysis 14, 6 (1994), 1101–1108.
9. GARERA, S., PROVOS, N., CHEW, M., AND RUBIN, A. D. A framework for detection and measurement of phishing attacks. In WORM ’07: Proceedings of the 2007 ACM workshop on Recurring malcode (New York, NY, USA, 2007), ACM, pp. 1–8.
10. ANDERSON, R., AND MOORE, T. The economics of information security. Science 314, 5799 (2006), 610–613.
11. GORDON, L. A., AND LOEB, M. P. The economics of information security investment. ACM Trans. Inf. Syst. Secur. 5, 4 (2002), 438–457
12. HERLEY, C., AND FLORˆENCIO, D. A profitless endeavor: phishing as tragedy of the commons. In NSPW ’08: Proceedings of the 2008 workshop on New security paradigms (New York, NY, USA, 2008), ACM, pp. 59–70.