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

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Vol. 3, Issue 4, April 2015

Detection of Websites Based on Phishing Websites Characteristics

Pallavi D. Dudhe, Prof. P.L. Ramteke

Student of M.E. in Computer Science and Engineering, HVPM's C.O.E.T., Amravati, India

Associate Professor and Head of IT Department, HVPM's C.O.E.T., Amravati, India

ABSTRACT: Phishing is a web-based attack that uses social engineering techniques to exploit internet users and acquire sensitive data. Most phishing attacks work by creating a fake version of the real site's web interface to gain the user's trust.. We applied different methods for detecting phishing using known as well as new features. In this we used the heuristic-based approach to handle phishing attacks, in this approached several website features are collected and used to identify the type of the website. The heuristic-based approach can recognize newly created fake websites in real-time. One intelligent approach based on genetic algorithm seems a potential solution that may effectively detect phishing websites with high accuracy and prevent it by blocking them.

KEYWORDS: phishing, heuristic-based,potential, genetic.

I. INTRODUCTION

The Internet is playing an increasingly significant role in today's commerce and business activities. Unfortunately, poor security on the Internet and large financial gains provide a strong motivation for attackers to perpetrate such seemingly low risk, yet high-return online scams. Email messages are not protected as they move across the Internet [1]. Often information being transmitted is valuable and sensitive such that effective protection mechanisms are desirable in order to prevent information from being manipulated or to protect confidential information from being revealed by unauthorized parties. The phishing attacker's trick users by employing different social engineering tactics such as threatening to suspend user accounts if they do not complete the account update process, provide other information to validate their accounts or some other reasons to get the users to visit their spoofed web pages. The word 'Phishing' initially emerged in 1990s. The early hackers often use 'ph' to replace 'f' to produce new words in the hacker's community, since they usually hack by phones. Phishing is a new word produced from 'fishing', it refers to the act that the attacker allure users to visit a faked Website by sending them faked e-mails (or instant messages), and stealthily get victim's personal information such as user name, password, and national security ID, etc [8]. These information then can be used for future target advertisements or even identity theft attacks (e.g., transfer money from victims' bank account).

II. PHISHINGWEBSITES DETECTION METHODOLOGY

A. Website characteristics related to phishing :

- IP address: Using an IP address in the domain name of the URL is an indicator someone is trying to access the personal information. This trick involves links that may begin with an IP address that most companies do not commonly use any more [2]. In the frequency analysis conducted earlier, 20% of the data contains ''IP'' address and all of them are associated with phishy websites. An IP address is like http://

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

91.121.10.211/~chems/webscr/verifySometimes the IP address is transformed to hexadecimal like http://0x58.0xCC.0xCA.0x62.

- Long URL: Phishers hide the suspicious part of the URL to redirect information's submitted by users or redirect the uploaded page to a suspicious domain. Scientifically, there is no standard reliable length that differentiates between phishing URLs and legitimate ones. (Mohammad et al. 2012) suggested when the URL length is greater than 54 characters the URL can be considered phishy.
- URL's having @ symbol: The ''@'' symbol leads the browser to ignore everything prior it and redirects the user to the link typed after it.
- Sub-domains: Another technique used by phishers to scam users is by adding a subdomain to the URL so users may believe they are dealing with an authentic website. An example: http://www.paypal.it.asce ndancethe atrearts.co.uk.
- Fake HTTPs protocol/SSL final: The existence of HTTPs protocol every time sensitive information is being transferred reflects that the user is certainly connected with an honest website. However, phishers may use a fake HTTPs protocol so that users may be deceived.
- DNS record: An empty or missing DNS record of a website is classified as phishy. Phishers aim to acquire sensitive information as fast as possible since the phishing webpage often lasts for short period of time and the URL is not valid any more. DNS record provides information about the domain that is still a live at the moment, while the deleted domains are not available on the DNS record [3] [4].
- Request URL: A webpage usually consists of text and some objects such as images and videos. Typically, these objects are loaded into the webpage from the same server of the webpage. If the objects are loaded from a domain other than the one typed in the URL address bar, the webpage is potentially suspicious.
- URL of anchor: Similar to the URL feature, but here the links within the webpage may point to a domain different from the domain typed in the URL address bar that a legitimate webpage has a rank less than or equal to 150,000.
- Redirect page: When users clicks on a link they may be unaware that he's redirected to a suspicious webpage. Redirection is commonly used by phishers to hide the real link and lures the users to submit their information to a fake site [5].
- Hiding the links: Phishers often hide the suspicious link by showing a fake link on the status bar of the browser or by hiding the status bar itself. This can be achieved by tracking the mouse cursor and once the user arrives to the suspicious link the status bar content is changed.
- Email: There is a function on PHP called mail or email and it take our information which we enter in the forms like "MasterCard number, etc. "and send them when we press the pay button throw e-mail to the phishers e-mail. Phisher can insert PHP code inside Html code and use this function to send our information [10] [11].
- iframe:It is HTML tag code and used to embedding another webpage into current webpage. It creates a frame or window on a webpage so that another page can load inside this frame. Phishers use the iframe and make it invisible i.e. without frame borders, when the user goes to website, he/she cannot know that there is another page is also loading in the iframe window. It is a big problem which all people do not know it, it is like small website open in current webpage for example: we can open www.google.com in my page www.mona.com by using iframe so when the people enter our website they will see the secured website is opened but it is not in the page it open throw iframe . Example: http://www.phisher.com/index.php?search="><iframe src=http://google.com ></iframe> // Replace http://google.com by the phishing page [6].
- Script:It is PHP files and there is some of phisher used scripts to send personal information or PC information to them, and some scripts send viruses or load from external websites. Scripts tag use to put any external file in the page like jquery or CSS and if it is with start and end tag, it is legal because this is

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

the correct and standard script tag.Example:<scripttype="text/javascript"src="includes/jscripts/jquery.min.js"></script>, it is now load file to make the page appearance good. When there are tags like this <script> and between them any codes (not links) it is suspicious tags because this script code is javascript or any other languages which may be used to send personal information or PC information to phishers. So if we find <script> tags and there end tags </script> it is a legal tags, otherwise it is a phishing character.

III.SYSTEM DESIGN AND IMPLEMENTATION

A. System overview:

Detect the phishing websites by checking the websites characteristics, we extract some phishing characteristics out of the all standards to evaluate the security of the websites, and check each websites name, if we find a phishing websites, we will block that websites and if we will find the websites as secure then we will open the home page of that websites.

B. System description and work flow:

- First:We enter the websites name for the detection of websites as phishing or not. Figure1 is the main window for program

Figure 1: Phishing detection program main window

- Second:Then based on the websites characteristics it will find out the types of websites either as safe or phishing websites. The characteristics used for detecting a websites are number of links, number of adds, java scripts and malware spyware data [9].

- Finally: Shows the window that indicates if this websites is secured or Phishing based on the above websites characteristics. The program workflow as follows in figure2. In this user enter the websites as input, then on the basis of characteristics that our system used , the url is detected either as safe or normal websites or phishing websites. If the websites is normal or safe then it will open the home page of that particular websites. And if the websites is detected as phishing websites then our system will blocked it for safety purpose.

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

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Vol. 3, Issue 4, April 2015

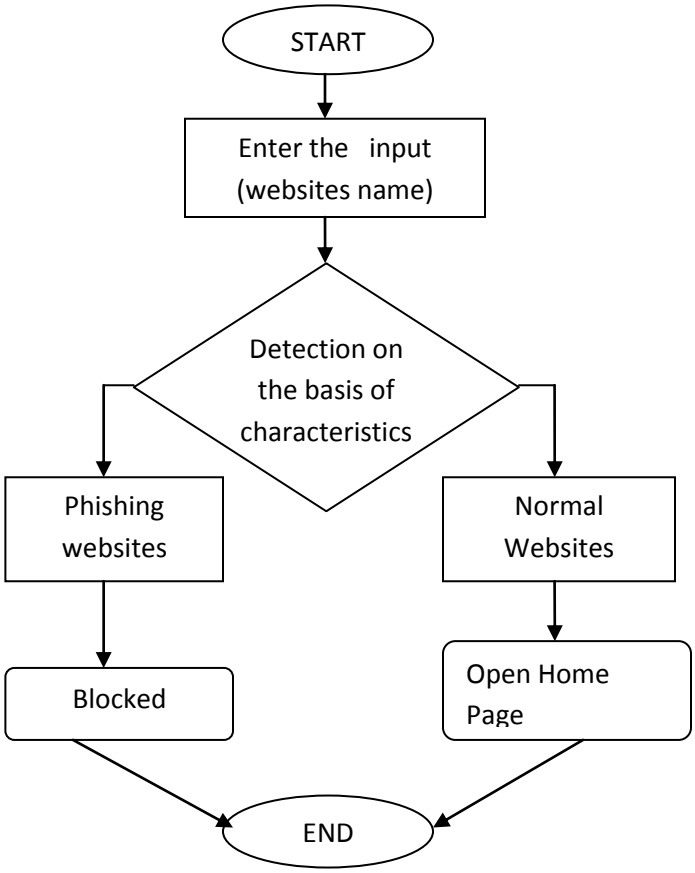

Figure 2: Phishing detection program flow chart

IV. TESTING AND RESULTS

A. Results screenshots:

The First window:The websites is safe websites detected on the basis of websites characteristics. Figure 3 is an example of testing results.

Figure 3: The program checking result window for safe websites

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

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Vol. 3, Issue 4, April 2015

The Second window:The websites is phishing websites detected on the basis of websites characteristics. As the characteristics that we are used for detecting the websites shows that the site will be phishing as shown below. As the external links in that websites is 123, number of adds on that sites is 246, java scripts available in that sites are 369 and malware and spyware data is also 369 on that websites, all these characteristics show that the input website is the phishing site Figure 4 is an example of testing results.

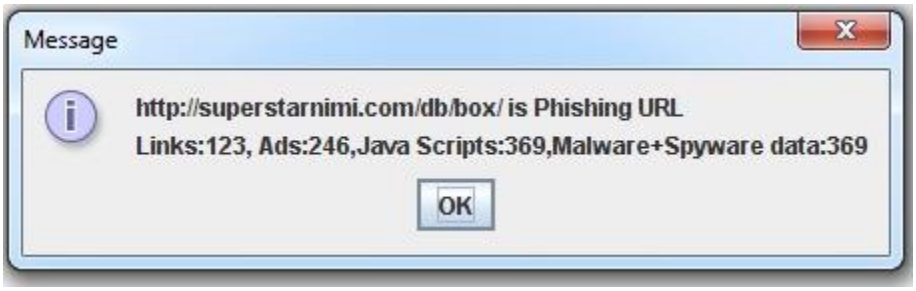

Figure 4: The program checking result window for phishing websites

As the websites is detected as phishing or unsafe then our program will blocked it for safety purpose.Figure 5 is an example of testing results.

Figure 5: websites blocked message window

As our proposed system detected the websites as phishing then for testing purpose as this website is actually phishing or not we enter the url (name of websites) in the browser then it will gives you the following message window for alert.

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

International Journal of Innovative Research in Computer and Communication Engineering
(An ISO 3297: 2007 Certified Organization)
Vol. 3, Issue 4, April 2015

Figure 6: Alert message

On this basis it will show that the websites that our system detected as phishing is actually phishing websites as shown in the above figure 6.

B. Comparison with Various characteristics

Following tables gives you idea about the risk involved in websites characteristics

Phishing characteristics	Phishing characteristics risk
Https	Medium
Images	Low
Suspicious url	High
Email	High
iframe	Low
Java scripts	High

V. CONCLUSION

Detecting the phishing websites is one of the crucial problems facing the internet community because of its high impact on the daily online transactions performed. There is no doubt that phishing, as a phenomenon, is both highly successful and generally difficult to detect and prevent in a reasonable amount of time. There are variety of phishing detection technique are available but our approach that is heuristic approaches gives the better result than other approaches. It detected the websites based on various characteristics and in minimum time period. Heuristics can detect phishing attacks as soon as they are launched, without the need to wait for blacklists to be updated. To make the Genetic Algorithm applicable for generating rules for large data sets it should be made scalable. There is still a big gap towards finding an optimum anti-phishing solution against phishes.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

REFERENCES

[1] L. A. T. Nguyen, B. L. To, H. K. Nguyen, and M. H. Nguye "Detecting phishing web sites: A heuristic URL-based approach," in Advanced Technologies for communications (ATC),International Conference on, pp. 597- 602,2013.
[2] Abdelhamid, N., Ayesh, A., & Thabtah, F. Associative classification mining for website phishing classification. In Proceedings of the ICAI(pp. 687–695), USA, 2013.
[3] Aaron, G., & Manning, R.APWG phishing reports 2012.
[4] Sophie Gastellier-Prevost, etc., "Decisive heuristics to differentiate legitimate from phishing sites", *Proceedings of IEEE*, 2011.
[5] J. Zdziarski, W Yang, and P. Judge, spam conference, Phishing activity trend report 1st half 2011
[6] J.shreeram, M.subam, P.shanthi, K.manjula. Sastra University Kumbakanam "Anti phishing detection of phishing attacks using genetic algorithm".Retrieved on October 8, 2010.
[7] Mather Aburrous, M.A. Hossain, Keshav Dahal, Fadi Thabtah "Prediction phishing websites using classification mining techniques with experimental case studies" in proceedings of Seventh International Conference on Information Technology, Las Vegas, NV, pages 176-181, 2010.
[8] Xiang, G., Pendleton, B. A., Hong, J. I., and Rose, C. P.A hierarchical adaptive probabilistic approach for zero hour phish detection. In Proceedings of the 15th European Symposium on Research in Computer Security (ESORICS'10). 268–285, 2010.
[9] T. Venkat Narayana Rao et al., " / Genetic Algorithms and Programming-An Evolutionary Methodology", *International Journal of Computer Science and Information Technologies*, Vol. 1 (5), 427-437, 2010
[10] V. Shreeram, etc., "Anti-phishing detection of phishing attacks using genetic algorithm", Proceedings of ICCCCT'10, pp. 447-450, 2010
[11] Maher Aburrous, M. A. Hossain, Keshav Dahal, Fadi Thabtah, "Predicting Phishing Websites Using Classification Mining Techniques with Experimental Case Studies, "Information Technology: New Generations, Third International Conference on, pp. 176-181, 2010.

BIOGRAPHY

Miss Pallavi D. Dudhe has Completed Degree in B. Tech (Computer Science and Engineering) from Government College of Engineering, Amravati, Maharashtra, India. She is pursuing M.E. Second Year Computer Science and Engineering in HVPM COET, Amravati, Maharashtra, India.

Prof. P. L. Ramteke is working as an Associate Professor, Department of Information Technology, HVPM COET, Amravati, Maharashtra, India. He received Master of Engineering (M.E) degree in Computer Science and Engineering from SGBAU, Amravati, MS, India. He is also completed his M.Phil in Computer Science Stream.