Scientific Journal of Impact Factor(SJIF): 3.134

e-ISSN(O): 2348-4470
p-ISSN(P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2,Issue 11, November -2015

Website Phishing Monitoring

Jaydeep Solanki

Computer Department, Darshan Engineering College Rajkot, India

Abstract- *Internet has become a useful part of our regular daily social and financial activities. People are heavily depends on internet and online activities such as online shopping, online Banking, online Booking, online Recharge and many more. Phishing is a form of web threat, phisher create the replica of original website and illegally try to get Victim's personal information like user name, password, credit card details, SSN number and use it for own benefit. A Non regular user cannot identify whether website is phished or legitimate. There is no any single solution to stop this fraudulent activity. This paper propose the model which identify the phishing site. We first extract the features which clearly differentiate that whether website are phished or legitimate. Then we apply this features to machine learning techniques it will identify website are phished or legitimate. In this way it will help towards our society.*

Keywords – *Phishing, Machine Learning, URL, Classifier, Phishing Site*

I. INTRODUCTION

Due to computer crimes (cyber-crime) getting popular and effective every day because of the lack of the security knowledge in common public. Resultantly it will cause very serious damage in many areas by Identity theft, Loss of private information, customer's confidential information theft in E commerce and online banking.

Phishing is one of this cyber threat in which attacker try to get victims confidential information directly from him/herself by sending them different types of notification such as Threat message or information updating request. There are many different techniques for phishing such as phishing by email, mobile phishing (SMS, instant messages) and website phishing [1]. Usually phisher make use of email or website to attack the unsuspecting users. This paper is focusing on website phishing, in website phishing attacker target unsuspecting victims. Attacker will send fraud notification of threat message or information updating message with link which will redirect to the website made by attacker which looks same as the legitimate website. Attacker simply make replica of the website to lure the victims and make them to give their confidential information.

According to the Anti Phishing Working Group (APWG) trends reports [2] it is clear that phishing attacks are increasing very rapidly. APWG report statistics shows, June 23 2014 phishing sites leaped by 10.7% in Q4 of 2013. And according to August 29 2014 128,378 phishing sites were observed which is 2nd highest number of phishing attacks after 164,032 seen in 1st quarter of 2012. More from APWG global survey for 2nd half of 2014 Top ten companies are targeted constantly and sometimes more than 1,000 time per month. Which resultantly cause damage of billions of money every year.

Basically two approaches are employed in website phishing identification [3]. The first one is list based approach in which Blacklist and Whitelist In this approach list will be used to identify phishy website. This approach needs up to date database of Blacklist and Whitelist. Second is heuristic-based Method where several features are collected from the website and used it to identify it either as phishy or legitimate.

In this paper we have proposed the website phishing detection system. The rest of the paper is organized as follows: Section II provides the literature review of related studies. Section III gives the detail about phishing website featureswhich will be used to distinguish between phishy and legitimate. Section IV provides proposed system for detecting phishing sites. Section V conclusion of work and its future direction.

II. RELATED WORK

As cyber criminals is taking more and more interest in phishing attacks because phishing is comparatively more easy and effective weapon to use against any common person. Hence Researchers have proposed many anti-phishing methods. In this section, we will discuss some of the related works.

CATINA is one of the very popular content based anti-phishing technique proposed by Zhang et al [4]. In CATINA term frequency-inverse document frequency (TF-IDF) calculated and keywords will be extracted from the webpage content and generates the lexical signature. Which will later use to perform google search and the result will be used for the classification of the websites. However, CANTINA performance will be influenced by the website language and also fails to track the brand names as keywords.

Ee Hung Chang et al proposed content based approach using website logo[5]. This method captures the screenshot of the webpage and extract its logo by logo segmentation process. Then the website logo will be used for identification process using google image database and google image search facility. This work require the updated image database.

Many feature based methods are proposed which uses some website features to clearly differentiate between legitimate and phishy. Rami M. Mohammad et al author used 17 features to identify phishing website using self-structuring neural network to classify legitimate and phishing website[7].

III. PHISING WEBSITE FEATURES

There are several features distinguish phishing websitesfrom legitimate ones. Binary valuefeatures hold either ‘‘phishy’’ or ‘‘legitimate’’ since theexistence or lack of the feature within the website determinesthe value assigned to that feature and later it will be used it identify the phishy website.

Based on the studies basically there are 27 features which are classified in 6 criteria which is listed below.

Table 1. Website Phishing Criteria

Criteria	Phishing Indicator
URL and domain identity	Using IP address
	Request URL
	URL of anchor
	DNS record
	Abnormal URL
Security and encryption	SSL certificate
	Certification authority
	Abnormal cookie
	Distinguished names certificate (DN)
Source code and JavaScript	Redirect pages
	Straddling attack
	Pharming attack
	Using on MouseOver
	Server form handler
Page style and contents	Spelling errors
	Copying website
	‘‘Submit’’ button
	Using pop-ups windows
	Disabling right click
Web address bar	Long URL address
	Replacing similar characters for URL
	Adding prefix or suffix
	Using the @ symbol to confuse
	Using hexadecimal character codes
Social human factor	Much emphasis on security and response
	Generic salutation
	Buying time to access accounts

IV. PROPOSED SYSTEM

In this section we describe system to detect phishy website. We will use Non Content based approach in which we will use website URL as a input in our model.

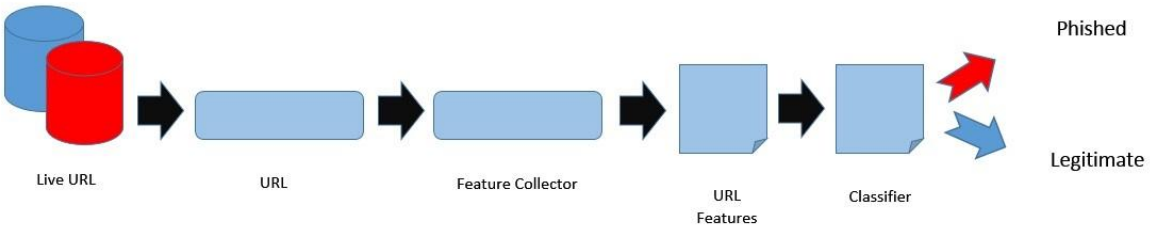

Figure 1. Proposed System

Step 1) Live URL Input: - Live requested URLs will be feed as input to the system. Phishing URL's dataset will be collected from PhishTank API, PhishTank is a community site where anyone can submit, verify, and share phishing URLs. And legitimate URLs dataset will be collected manually.

Step 2) Feature Collector: - Feature collected by using .NET framework. Phishers can use URL's to hide the doubtful part in the address bar. The feature extractor will extract various features from the URL like.

1. Lexical Features: Length of URL, No. of slashes in URL, The number of dots, special character it contains
2. Domain based Features: Age of Domain, DNS Record
3. Host Features: of URL contains, IP address

Step 3) Classification: - After extracting all this features the file will be provided as input to the classifier which will classify it to Phishy or Legitimate. Here classifier will be the Machine Learning algorithm which will be trained first by the testing URL dataset.

V. CONCLUSION

This work system of phishing website prediction as a classification task and demonstrates the machine learning approach for predicting whether the given website is legitimate website or phishing. Features have been extracted from URL dataset.

It is hoped that more interesting results will follow on further exploration of data.

REFERENCE

[1] Rani S. K, 2D. Kavitha M. E., International Journal of Technical Research and Applications e-ISSN: 2320-8163, www.ijtra.com Volume 3, Issue 2, PP. 45-51, Mar-Apr 2015.
[2] http://www.antiphishing.org/resources/apwg-reports/ [online]
[3] Rami M. Mohammad,FadiThabtah, Lee McCluskey, University of Huddersfield Repository WORLDCOMP. World Congress in Computer Science, Computer Engineering, and Applied Computing, Las Vegas, Nevada, USA, ISBN 1601322461, pp. 682-686., 2013.
[4] Y. Zhang, J. I. Hong, and L. F. Cranor, CANTINA: A content-basedapproach to detecting phishing web sites. roceedings of the 16th International Conference on World Wide Web, pages 639–648, 2007.
[5] Ee Hung Chang, Kang LengChiew, San Nah Sze, Wei King Tiong, ScienceDirect/ Computer and security, 7 August 2015.
[6] Radheshyam Panda, Rajesh Tiwari Dept.of CSE, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 3, ISSN: 2277 128X, CSVT University, Bhilai, India, March 2014.
[7] Rami M. Mohammad, FadiThabtah, Lee McCluskey, 17 April 2013 / Accepted: 10 September 2013 / Published online: 21 November 2013_ Springer-Verlag London, 2013.
[8] Ms. KrantiWanawe, Ms. SupriyaAwasare, Mrs. N. V. Puri, International Journal of Research in Advent Technology, Vol.2, No.3, E-ISSN: 2321-9637, March 2014.
[9] Ram B. Basnet, Andrew H. Sung, Quingzhong Liu, Institute for Complex Additive Systems Analysis (ICASA), New Mexico Tech, Socorro, NM 87801, USA.

