International Journal of Engineering & Technology, 7 (4.39) (2018) 147-149

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Design a Client-Side Application with Automatic Real-Time Phishing Detection Mechanism based on the White-Lists

Sumit Wadhwa[1*], Arun Pratap Srivastava[2], Shashank Awasthi[3], Mahesh Kumar Singh[4]

[1,2,3,4] G.L.Bajaj Institute of Technology & Management
**Corresponding author E-mail: apsvgi@gmail.com*

Abstract

Phishing is a problem involving not to be trusted emails and WebPages that trick unsuspecting users into willingly revealing their confidential information. Phishing is a cyber attack which involves a fake website mimicking the some real legitimate website. In this paper we design, implementation, and evaluation of phishing detection using HTML/CSS, PHP and store the credential details in session for the being time and can work with any authentication technologies which are based on exchange of credentials also discuss the detection of login phishing page that contains email and password, credential fields that provide personal/restricted content. Integrated security feature is one of the best solution for prevention a phishing attacks to secure web browser when-ever a phishing site is accessed by an internet user

Keywords: *Phishing detection, Domain Name, Heuristics, URL analysis, Login pages, White-list, Web security*

1. Introduction

In the phishing attacks, detection and prevention are the very big challenges as the phisher per-forms attacks to bypass the existing anti-phishing techniques. An educated and experienced user may still fall this attack. The attacker makes a fake yet similar webpage by copying or making a little change in the legitimate page for e.g- www.rediffmail.com to www.redifmail.com, so that an internet user will not be able to differentiate between the real and the phished one. For example, a system can be technically secure enough against password stealing, however uninformed end users if click on the Hypertext Transfer Protocol (HTTP) link may leak their passwords, which ultimately threatens the overall security of the system. There are many solution exist to detect phishing attack but no one bullet proof solution yet to present which detect all type of phishing attack. Generally, web browsers provide security against phishing attacks with the help of list-based solutions like applications which support client server architecture.

2. Literature Survey

Cao Y et al. [1]: Proposed one can gather a list of legitimate URLs. This method is recognized as white listing, and it is also a kind of list-based approach. An example of a white listing technique is the research proposed by the authors developed an automated technique that maintains and stores a white list at the client side.web sites, we decided to check this and so we compiled a list of target words which included many popular phishing targets, such as E bay and pay pal. In most of cases criminals make web pages by copying legitimate or make a little change. [1].
Deshmukh et al.[2]: Proposed a approach as cyber crime is technology based fault committed by technocrats. This paper deals with modification of cyber crime like Packet Sniffing, Salami Attack, Bot Networks and Tempest Attacks. It also contains real world cyber crime suitcases their situation and modus operandi. The worldwide malware, rate spam rate and phishing rate is rising speedily. And there is a latent shock of cyber crime on consumer trust, economics and production time. The contradict ways similar to Intrusion Detection, GPRS Security architecture and Agent Based Distributed Intrusion Detection System and prevention System are utilized for safety reason
Jo et.al in [3]: In this paper it was proposed an approach based on websites' identity claims. The system copy the human behavior of accessing the website, system study and identity website, assert, and figure the documentary significance between this claimed identity and other description in the website. Their phishing detection system then employs this textual significance as one of the sort for classification.
Tan et.al in [4]: Proposed an anti-phishing method to protect users against phishing attacks in the internet. The scope of this approach study focuses mainly on the detection of phishing websites with English content. In order to encourage users on whom the website claims to be, phishers usually place brand names in different parts of the URL. They oppressed this phishing pattern by conveying weights to words take out from the HTML content, based on their co-appearance at path, hostname and file names of URLs. These weights are then supplementary to their equivalent TF-IDF weights. The most likely words are particular and submitted to Yahoo Search to recover the highest frequency domain name amongst the top 30 search results. A WHOIS lookup is carry out to disclose the vendor behind the selected domain name. A phishing website can be easily illustrious if the vendor of query domain name be different from the owner of domain name returned by the search engine.

3. Phishing Detection Life Cycle

Our methodology is to build a client-side application with automatic real-time phishing detection mechanism based on white-list.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The application will inject fake credentials to login pages and check the response of the website. This approach is quite similar to the Phish Guard's [20] approach which was discussed in section 3 "literature view". Phishing login pages are designed to lure victims into willingly giving their credentials. However, Phishing websites has no information regarding the victim's real credentials

i. The phisher clones the content from the website of a legitimate company or a bank and generates a phishing website. The phisher tries to keep the visual similarity of the phishing website to the corresponding legitimate website to trick more users
ii. The phisher sends an email including the link of the phishing website to it to his victims. In the case of spike phishing, a mail is sent to individual targeted victims.
iii. When the victim opens the email, and visits the phishing website, the phishing web-site prompts the victim to insert private data, for example, if the phisher copycats the phishing website of a famous organization, then the users of organization are expected to willingly reveal their private credentials to the phishing website.
iv. The phisher receives private data of the victim via the phishing website and utilizes this data for financial or some other benefits which will be discussed in detail is in the background section.

4. New Proposed Scheme

Web pages are being fully tested and classified by the phishing identification module. The phishing identification module tests login pages by filling the login fields with fake credentials multiple times and based on the response, the page will be classified as phishing or legitimate. The phishing detection module is written in PHP, HTML/CSS.

(Algorithm)

Step 1: This method design a client-side application with automatic real-time phishing detection mechanism based on the white-lists.
Step 2: inject all the fake credentials (Login id & Password) to login pages and check the response of the website.
Step 3 Phishing login pages are designed to lure victims into willingly giving their credentials, if website has no information regarding the victim real credentials then expected outcome like phishing website show a failure message or redirect to another website.
Step 4: then user visits a random website with meaningful URL(Uniform Resource Locator) and DN(Domain Name) module to check it is present in white list or not?
Step 5: If domain name is found or available in white list then the website is legitimate.
Step 6: If no domain is found in the white-list, the URL and Domain analysis phase will check the following parameters of the URL
Number of Dots in the URL > 5
Special Character "@" in the URL.
Domain expiry date< 30 days
Domain creation date < 365/366 days
Step 7: once user enter their credentials like user id and password which is not in the form of MD5 are stored in session and cookie, because user unable to identify the exact difference between duplicate page or original page once user enter their details are stored in session and hacker can access their social media network or bank details as well.
Step 8: The Phishing identification module will inject n number of fake email id and password combinations.

(i) The phishing detection module will detect all the input text fields with type attribute like (email) or name attribute (email, user, username, Id and user id).
(ii) The module will inject the input text fields with fake credentials and wait till page loads.
(a) If the page loads with no password fields or blank then it is considered as phishing.
(b) If the page redirects to another website then it is considered as phishing.
(c) If the page reloads with an input text field of type password, then the application will inject more fake credentials for n number of times.
(d) If the password field still exits after n number of trials, then the page will be considered as legitimate

Step 9: Exit

5. Phishing Identification Module Results

The results and evaluation of our tool "SeleniumPhishGuard" will be discussed in this section. To ensure that our tool is language independent datasets collected from different languages websites Firefox will find and inject email and password in their corresponding fields. Then it will submit form by simulating "enter" key (See Figure 5.1).



Figure 5.1: Selenium filling Facebook login form

Web-driver will continue to inject email and password and submit the form for n number of times. After that it will check for the presence of the password field (See Figure 5.1). If the password still exists then the webpage will be considered as legitimate, else classified as phishing. Test results are shown in the output terminal as shown in Figure 5.2.



Figure5.2: Page response after form submission

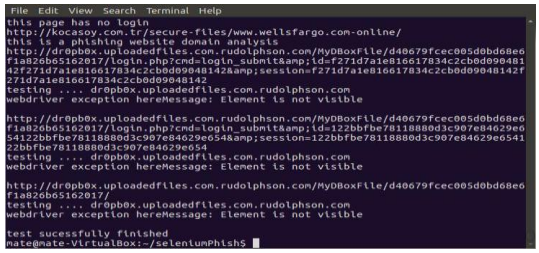


Figure 5.3: Test successfully finished

Real time data visualization was achieved by using Grafana. Grafana is a graphical tool integrated with our influx time series logging database to show results in real-time (See Figure 5.4).



Figure5.4: Grafana visualizing False positives and True Positives

Table 5.1: Heuristics weights

Heuristic	True positive	False positive	Effect	weight
Domain creation Date	85%	32%	53	5
Domain expiry date	23%	5%	18	2
'@' in URL	10%	0%	10	1
'-' in URL	15%	4%	11	1
Dots in URL	44%	5%	39	4
Domain inWHOIS	9%	7%	2	0

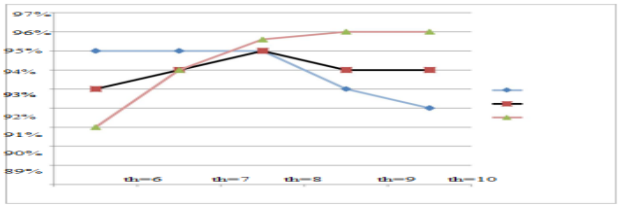


Figure 5.5: Graph showing the threshold effect on accuracy

To choose a value " " for threshold " h " function in equation no.1, the system was tested 5 times against 100 phishing URLS and 100 legitimate URLS dataset (3). The limits for the choice of threshold values was not chosen randomly. Domain creation date has a heuristic weight of 5 and it does not make sense to have a threshold value " " of 5 since the classification will be based only on this heuristic. Therefore, lowest thresh-old value in our test is 6, as the highest weight. The highest value for this test is 10, since its corresponding true positive rate is equal 93% same as using the phishing identification

5. Conclusions and Future Work

In this work we have developed a new method to detect phishing websites based on the URL of the website and all the relevant information are used in authentication process are stored in session or database to hack the records. The system has shown a 98% detection rate of phishing website because the white list used by phishing website returns the search of original websites or other websites that have back linked the original website but the test website's URL never appears in the search result. Thus making a 98% accurate detection because some of the records are stored in white list

In the future work, we can add more parameters like Google Page Rank, number of back links etc in order to increase the overall confidence towards phishing as well as non-phishing website.

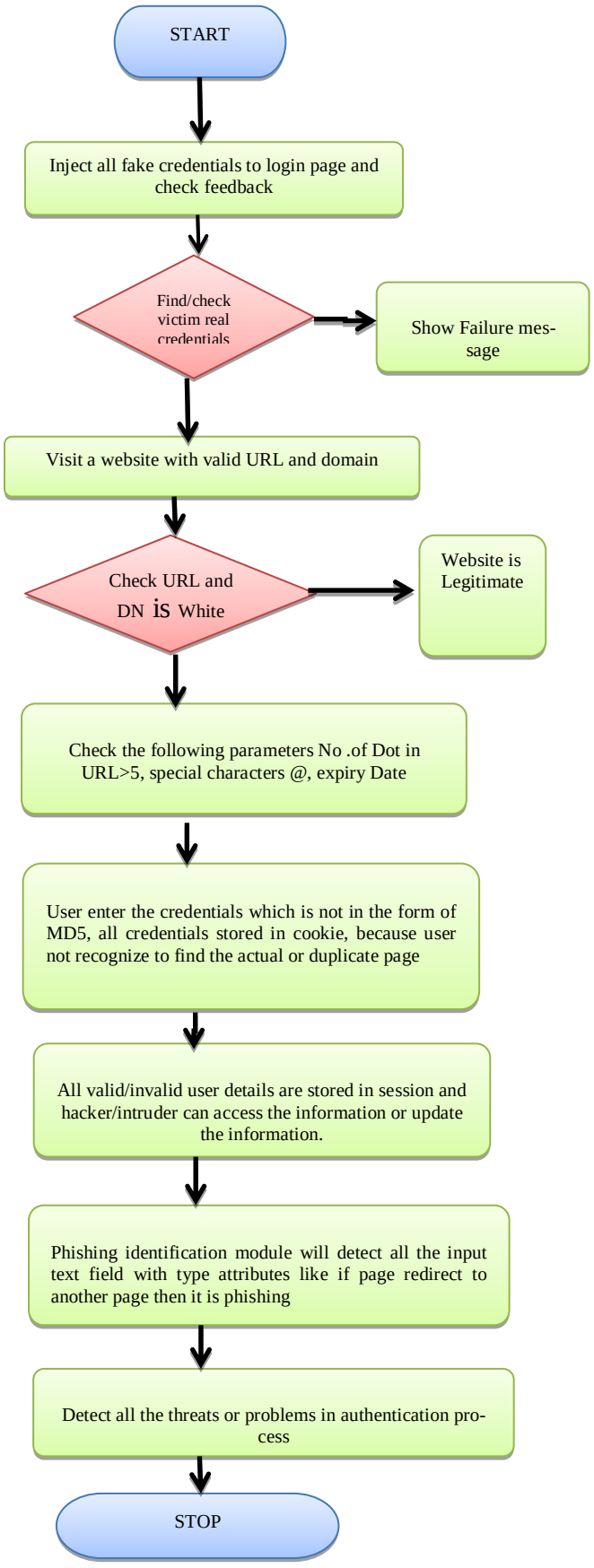
References

[1] Cao Y, Han W, Le Y, "Anti-phishing based on automated individual white-list," Proceedings of the 4th Workshop on Digital Identity Management., pp. 51e60,2008.

[2] Deshmukh, J.J. and Chaudhari, S.R., 2014. Cyber crime in indian scenario–a literature snapshot. International Journal of Conceptions on Computing and Information Technology, 2(2).Marc Parenthood, Patrick Rainier, Jacques Tissue.

[3] Jo, I., Jung, E.E. and Yeom, H.Y., 2010, August. You're Not Who You Claim to Be: Website Identity Check for Phishing Detection. In Computer Communications and Networks (ICCCN), 2010 Proceedings of 19th International phishtank.com/. [Accessed 01 July 2016]..

[4] Tan, C.L. and Chiew, K.L., 2014, December. Phishing website detection using URL-assisted brand name weighting system. In Intelligent Signal Processing and Communication Systems (ISPACS), 2014 International Symposium on (pp. 054-059)



Flow chart 5.1: Phishing Detection System