ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

A Framework to Protect Client Side Script Phishing Attack Using White-List

P.Malathi[#1], Dr.P.Vivekanandan[#2]

Assistant Professor, Department of Master of Computer Applications, Dhanalakshmi Srinivasan College of Engineering and Technology, Chennai, India [#1]

Professor of Eminence, Department of Chemical Engineering, A.C.Tech, Anna University, Chennai, Tamilnadu, India[#2]

ABSTRACT: Several anti-phishing solutions are having two major restrictions. The first one is the fast access time for the real environment and the second one is the high detection rate. Visual similarity and machine learning are having high detection rate but it has low access time. It is proposed in this method to prevent client side using white-list of the legitimate website accessed by the end user. The experimental results imply that the proposed system produces better detection rate and quick access time. In addition to that, the proposed methodology also aims at reducing various attacks namely zero-hour attack, embedded objects etc.

KEYWORDS: Phishing attack, Spoofed email, back door, Spyware DOS attack

I. INTRODUCTION

Phishing is a social attack that aims to exploit the vulnerabilities of the system processes caused by the end user. Phishing is the word extracted from "Phreaking" and "Fishing". Phreaking means making phone calls for free. Fishing means using bait to lure the target. In the phishing website, the phisher includes official logos from real organizations and other identifying information from the authentic website. Frequently used attack methods are sending e-mail and phishing websites or links. For example, the phishing site may ask the user to re-enter the password many times and run the upgrading services to capture the user's new credentials like password, account number etc.,

The process of phishing attacks are:

1) Phisher set up a phishing website which looks exactly like the legitimate website.
2) Phisher sends a large number of spoofed e-mails to the target user. Without any knowledge, the user may open the spoofed hyperlink in the e-mail and enter the information whatever enquired by the phisher.
3) Phisher steals the personal information and performs money transfer from user to the phisher.
4) The phisher can run client-side script applications to steal the user's personal information from the system unknowingly.

1.1 Types of Phishing Attacks

- ✓ **Content-Injection Phishing:** In this attack, the phisher replaces some of the contexts of the legitimate website with the false context. Therefore users are mislead to access the non-legitimate website and feed their valuable information.
- ✓ **Session Hijacking:** When the user signs in a session for a transaction and all bonafide credentials will be hijacked by the malicious software. Further transactions and other unauthorized actions will be carried out by the malicious software without the user's knowledge.
- ✓ **Deceptive Phishing:** It is originally referred to an account theft using instant messaging and e-mail messages.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

- ✓ **Web Trojan:** It concentrates on Popups. The Pop-ups are invisible when the user makes an attempt to log in. In that time the user's credentials are forwarded to the phisher.
- ✓ **Spear phishing:** This phishing attempts aimed at specific individuals or companies. The attackers may gather personal information about their target.
- ✓ **System Reconfiguration:** In this attack, the user system settings are altered so that the legitimate URL is changed which leads to system reconfiguration. Ex: the URL "uber.com" is changed as "ubre.com".
- ✓ **Malware-Based Phishing:** This attack is due to the malicious software running on user PCs. Malware is defined as a downloadable file from a website.
- ✓ **Data Theft Attack:** During this attack, the malicious codes which are running in the user's system can directly steal the confidential data in the system.
- ✓ **Search Engine Phishing Attack:** In recent years, some of the websites were created with fake offers. Also, these websites are marked by various search engines as legitimate websites which enable the user to trust and feed their information in the non-legitimate websites
- ✓ **Host file Poisoning:** It resolves domain names to Internet Protocol (IP) addresses. It resides on the windows and overrides DNS name resolution. Henceforth, the user is pushed to a phished website. Phishers extract the hosts' file and steal the user's information through a file address which means the URL contains a duplicate address.
- ✓ **Pharming Attack:** The phishers will alter the host file or DNS name so that the user will not be aware of the phished website.
- ✓ **Man-in-the-Middle Phishing:** It prevails on the data communication exists in the website The phisher will record the user's information by not disturbing the current communications or transactions. Afterward, the phisher may misuse the recorded credential information.

Consider a phishing attack with an example. The user is browsing a specified web page. The page redirects the user to click certain operation. The user clicks the link and does not know the page is phished or not. The phished link interacts the user to withstand in the page until it grabs the information from the specified user. The phishing is held in the following scenario.

- If the user clicks a particular link at a certain level, then the user may not be able to return back to the home page.
- The user will be diverted to other links
- User forgets the browsing history.
- The phisher extracts the details of the user when the link crosses certain level.

The various notable attacks that affect internet users are listed here

1.2 Zero-Hour Phishing Attack

This is one type of hole in anti-phishing technique which is unknown to the user. The attacker uses this security hole and hacks the information before the user awareness.

1.3 Embedded Objects

A proxy web page is developed and in that the attacker hides the address bar by projecting images or script and makes to believe, it is an original website.

1.4 DNS Poisoning Attack

Normally the DNS will be overcrowded with different websites. In order to reduce the data traffic in the DNS, the phisher will transfer the user control to the phished website from the legitimate website. In this attack, the users will work on the phished website desired by the phishers. For example, an attacker spoofs the IP address DNS entries for a target website on a given DNS server and replaces them with the IP address of the server under their control. Then the attacker creates the malicious content files on the target server. At that time when the user accesses the site, the poisoned DNS diverts the user to accept malicious content coming from phishing server and unknowingly downloads the malicious content.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

II. RELATED WORK

The author Sophie Gastellier [3] proposed a DNS based approach to detect the pharming attacks. Both the IP address check and the webpage content validation are performed to evaluate the website. whenever the user accesses the website, the DNS request is sent to both the default server and the third-party server. The user may choose any of the existing free name servers like OpenDNS, Google DNS as third-party server. The website is considered as legitimate if the IP address returned by the default server is also present in the list returned by the third-party server. Otherwise, the source of the web page will be extracted and the content will be examined to classify whether the site is legitimate or not.

The authors Debra L Cook et al. [4] discussed a new phishing filter and their advantages over the existing filters. This filter is named as phishwish and it does not need any training or white-list or blacklists. Furthermore, it is simple to configure and requires only 11 rules to determine the integrity of an incoming email. It compares the performance of phishwish to Spam Assassin and Google's browser-based phishing filter. The results indicate that phishwish outperforms these filters and identifies zero-day phishing attacks that were undetected by the existing filters.

A layered anti-phishing approach CANTINA+ was proposed [5] to catch the frequently growing new phishing attacks and control the false positives through the filtering algorithm. The proposed approach is a feature-based and employs hash-based filtering which abuses HTML DOM, search engines and third party services with machine learning techniques to detect phish. The proposed method uses two filters to minimize false positives and increase the runtime. A near duplicate phishing detector is used to catch the highly similar phish through hash-based filtering and the login form filter, which directly classifies the website without feature extraction. The authors conclude that the proposed approach achieves a better accuracy of 92% in unique testing. The proposed method cannot detect the XSS attack, host poisoning phishing attack, and content-based phishing attack.

An effective anti-phishing approach was proposed by using Automated Individual White-List (AIWL)[2]. This approach has been routinely maintained a white-list of the user's all frequent websites based on the naïve bayes classifier. AIWL intimate the user to the phishing attack. AIWL efficiently protected the user from the pharming attack. This approach effectively identifying phishing attacks and pharming attack.

III. PROPOSED METHOD

The proposed system comprises of the five phases, namely, (i) white-list maintenance, hyperlink evaluation, DNS Lookup analysis, and domain analysis. whenever the user accesses the website, the detection algorithm will check for its existence in the white-list. If the domain name exists in the white-list, the user will be allowed to access the website after validating the DNS and IP address. If there is access to the website which is not present in the white-list, the hyperlink will be extracted and validated. The detection algorithm evaluates the extracted links and classify the access as phishing or genuine.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

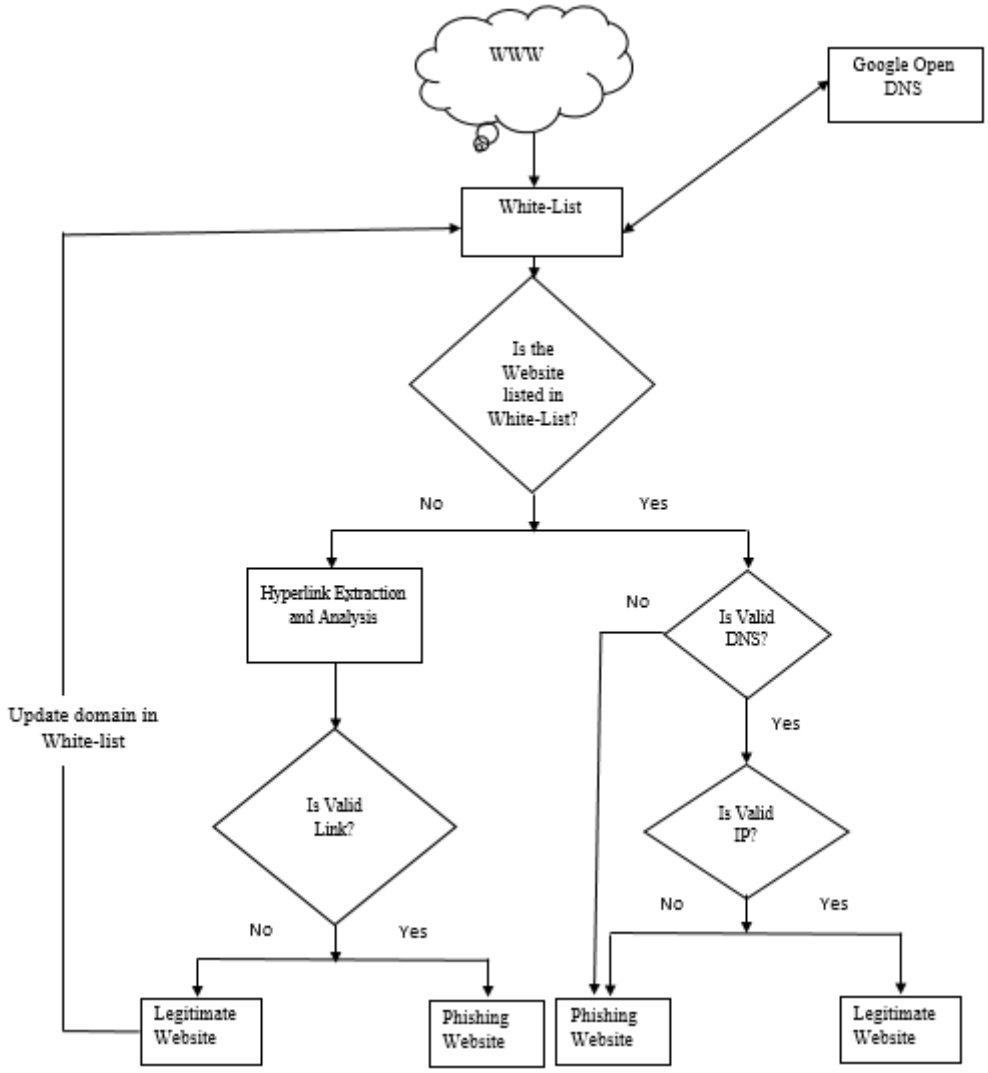

Figure 1: The Mechanism for Phishing Website Detection

3.1 White-List

The white-list has been prepared by referring the domain names from the Google OpenDNS server. Phishtank is a free online community site provided by OpenDNS where the users can submit, verify and validate any suspicious phishing site. The OpenDNS analyzes the phishtank data by identifying the Autonomous System Number (ASN) of the organization that blocks the suspicious IP address. Since the domain names of genuine sites are already maintained in the list, the detection time is low for evaluating the frequently accessed sites. The domain name and the relevant IP address are maintained in the White-list. Initially, the white-list is empty. If the domain name does not exist in the list, it will be further processed by extracting the hyperlinks.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

3.2 Domain Analysis

The DNS record consists of the following records.

- Address (A) records
- CNAME records,
- Name Server (NS) records
- Mail Exchange (MX) records.

Domain analysis checks whether the domain of the current website exists in the list or not. If it exists,then the control goes to the DNS Lookup analysis.

3.3 DNS Lookup Analysis

DNS lookup is the collection of various domain names that are linked with an IP address. The pointer records (PTR record) are used to resolve the IP address to the relevant name of the domain. This is similar to look up a phone number in a phone book. Interconnected computers, servers, and smart-phones need to know how to translate the email addresses and domain names used by the people into meaningful numerical addresses. A DNS lookup performs this function. The DNS lookup analysis is used to check the IP address matching. If the IP address is matched then the website is legitimate. Otherwise, the website is phished.

3.4 Hyperlink Extraction and Analysis

The hyperlink information are extracted from the page source of a particular web page using DOM object properties such as href, src, anchor, img, script and link tags. The phisher users have limited hyperlinks but the legitimate user would have several links. The examples are banking sites and organization sites. The motivation of hyperlink extraction is that the phishing page is the copies of the content of the legitimate information.

3.5 Detection Algorithm

The detection algorithm evaluates the website and classifies it. Detection algorithm takes the decision based on the parameters of hyperlinks. The website is classified as phishing if the number of hyperlinks in that site is zero or if the hyperlink contains the special character '#'. Otherwise, it is concluded as a genuine site. The detection algorithm is given below.

1. Calculate hyperlinks[n]
2. If hyperlink[n]=0 then
3. Print “WebPage Is Phishing”
4. End If
5. For hyperlink[n] > 0 then
6. If hyperlink[i]=”#”
7. Print “WebPage Is Phishing”
8. Else
9. Print “WebPage Is Legitimate”
10. End If
11. End For

IV. RESULTS AND DISCUSSIONS

The proposed approach protect the web users from a web browser plug-in for accessing websites in the client side. When the user accesses any website through the web browser the proposed detection approach classified as the particular website is phishing or legitimate. The below figure 4 shows the white-list modification in ten days. From this, it is observed that if the number days increased the white-list modification is decreased gradually.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

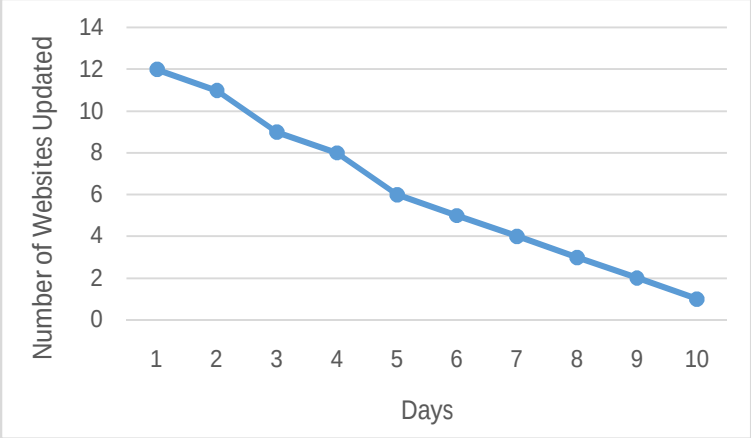

Figure 2: Monitoring the White-list

	Classified as Phishing	Classified as Legitimate
Phishing	TN	FN
Legitimate	FP	TP

Table I: Classification of Confusion Matrix

In any binary classification problem, where the goal is to detect phishing instances in a dataset with a mixture of phishing and legitimate instances, only four classification possibilities exist. The confusion matrix is presented in Table I. In Table I, the following are given. TN(True Negative) is the number of phishing instances that are correctly classified as phishing, FP (False Positive) is the number of legitimate instances that are incorrectly classified as phishing, FN (False Negative) is the number of phishing instances that are incorrectly classified as legitimate, and TP (True Positive)is the number of legitimate instances that are correctly classified as legitimate. The performance of a phishing web page detection system can be evaluated using the following metrics.

$$\text{True Positive (TP)} = \frac{Number\ of\ correctly\ classified\ genuine\ sites}{Total\ number\ of\ genuine\ sites}$$

$$\text{True Negative (TN)} = \frac{Number\ of\ correctly\ classified\ phished\ sites}{Total\ number\ of\ phished\ sites}$$

$$\text{False Positive (FP)} = \frac{Number\ of\ genuine\ sites\ misclassified\ as\ phished\ site}{Total\ number\ of\ genuine\ sites}$$

$$\text{False Negative (FN)} = \frac{Number\ of\ phishing\ sites\ misclassified\ as\ genuine\ site}{Total\ number\ of\ phished\ sites}$$

$$\text{Accuracy} = \frac{Number\ of\ Correctly\ classified\ sites}{Total\ number\ of\ sites}$$

V. CONCLUSION

The proposed approach prevents the client side phishing attack using white-list. This approach is used to check the legitimacy of a webpage using hyperlinks and detection algorithm. The proposed white-list framework reduces the false positive and increases the accessing time of the website. At the same time, it detects the DNS poisoning attack, embedded objects attack, zero-hour attack.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

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

Visit: www.ijirset.com

Vol. 7, Issue 3, March 2018

REFERENCES

1. Anti-Phishing Working Group (APWG), "Phishing activity trends report — second half 2010," http://apwg.org/reports/apwg report h2 2010.pdf, 2010, accessed December 2011. 28http://www.buhooth.ae/
2. Y. Cao, W. Han, and Y. Le, "Anti-phishing based on automated individual white-list," in DIM '08: Proceedings of the 4th ACM workshop on Digital identity management. New York, NY, USA: ACM, 2008, pp. 51–60.
3. Sophie Gastellier-Prevost, Gustavo Gonzalez Granadillo & Maryline Laurent 2011, 'A dual approach to detect pharming attacks at the client-side', IFIP International Conference on New Technologies, Mobility and Security, IEEE, pp. 1-5.
4. D. L. Cook, V. K. Gurbani, and M. Daniluk, "Phishwish: A stateless phishing filter using minimal rules," in Financial Cryptography and Data Security, G. Tsudik, Ed. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 182–186.
5. Guang Xiang, Jason Hong, Carolyn P Rose & Lorrie Cranor 2011, 'CANTINA+: A Feature-rich Machine Learning Framework for Detecting Phishing Web Sites', ACM Transactions on Information and System Security (TISSEC), vol. 14, no. 2, pp. 1-32.
6. Anti-Phishing Working Group (APWG), "Phishing activity trends report — first half 2011," http://apwg.org/reports/apwg trends report h1 2011.pdf, 2011, accessed December 2011.
7. Anti-Phishing Working Group (APWG), "Phishing activity trends report — second half 2011," http://apwg.org/reports/apwg trends report h2 2011.pdf, 2011, accessed July 2012.