IOSR Journal of Engineering (IOSR JEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
PP 06-08

Black Phish: A Novel Approach for Anti-Phishing

Ansari Aadil[1], Sachin Aware[1], Saud Mahevi[2]
[1](*Department of Comp. Engg, MMANTC, Mansoora, Malegaon, India)*
[2](*Department of Civil. Engg, MMANTC, Mansoora, Malegaon, India)*

Abstract: *In recent years, many researches have been going on topic of anti-phishing, phishing theft has become increasingly epidemic. Phishing attacks have become more profitable, especially with the popularity of e-banking and mobile payment. Exploring efficient and practical anti-phishing technology is particularly necessary and important. In this paper, by examining the resources (CSS, JS, and image files) request characteristics of phishing sites, we proposed BlackPhish, a novel anti-phishing method. BlackPhish as an enhanced blacklist technology, can detect not only phishes in blacklist, but also phishing which are emerging.*

Keywords:*phishing sites, phishing detection, anti-phishing, epidemic, blacklist.*

I. Introduction

A Computer user is often mislead to enter personal information at a fake website which appears to be legitimate whose looks are almost same to the original one, which is typically spread by email, Instant Messaging (IM), Quick Response (QR) code or other communication channels. This is called as Phishing. The U.S Federal Bureau of Investigation claimed cyber criminals successfully stole $3.5 billion from businesses during 24 months [1].

At present, the first thing is to identify phishing sites, anti-phishing mainly depends on technical countermeasures. Once the website have been identified it should be blacklisted. Based on the blacklist, mainstream browsers and security software prevent them from loading. The efficient and timely data update is an important prerequisite for the effectiveness of blacklist method.

In this paper, we propose a new novel anti-phishing method based on statistics of the reported phishing data, which detects phishes from a new perspective -- mining brand resource request relations. The proposed method is complement to existing methods and efficient, easy to implement and very effective.

II. Related Work

Anti-phishing can be categorized into three types as phishing detection, phishing disposal, and phishing discovery methods [2-4]. The Initial step is to learn a model to recognize phishing sites via extracted statistical features. The second is blacklisting, the way to dispose the phishing sites. The third is data mining, to find phishing sites actively. Statistical recognition as the conventional phishing detection method, the core is to identify the potential patterns of phishing sites by learnt models, which are trained with dissimilar features, such as Uniform Resource Locator (URL), visual information, page content and third-party services. CANTINA+ is a feature rich method, and it uses URL, Web page structure and third -party services features [2].

Blacklisting technology is common phishing disposal technology. In 2009, Sheng et al. studied the effectiveness of phishing blacklists. They tested the blacklists on a set of 15,000 and more legitimate URLs for false positives, finding no instance of mislabeling. They also show that blacklists are not effective for zero-hour phishes, and the true positive is less than 20% [3]. Prakash et al. proposed predictive techniques, which discovered new phishing URLs by enumerating simple combinations of known phishing sites [4].

In order to achieve the effect of imitate fraudulently, brand resources references are common in phishing sites. This paper mines brand resources request relation, and proposes a new phishing discovery method -- Resources Request based anti-Phishing (BlackPhish).

III. Black phish

Today, the Web page is consist of number of component. Most of Web pages have many resources (Cascading Style Sheets (CSS), JavaScript (JS) files, images, etc). Because of maximum concurrent connections limitation to the same domain for browsers, brand sites usually run resource content on another domain, such as PayPal runs CSS, JS and image files on paypalobjects.com.

However, most phishing sites are calling brand resources explicitly. In some cases, the html page has not any resource links to the sites, but the local CSS file may contain the brand resource, such as .gif file. The more deeply hidden way of branding resource calls is through JavaScript programming.

We analyzed phishing data of Anti-phishing Alliance and PhishTank. The statistics shows phishing sites employ external brand CSS, JS, or image files. So mining the resource request relation will be an effective way to find emerging phishing sites. Algorithm 1 illustrates the procedure of BlackPhish.

Algorithm 1: Resource request base anti-phishing method—BlackPhish

Input:	
url:	suspicious URL to be detected .
associatedDomainSet:	the domain name, whose query request are triggered when a browser loads the url page.
blacklist:	phishing domain name set.
whitelist:	brand domain name set.

```
1: Extract the domain name domain(url) of url;
2: if domain (url) ϵ whitelist
3:   ret = 0;
4: else if domain(url) ϵ blacklist
5:   ret = 1;
6: else
7:   collect associated DomainSet when browser loads url page;
8:   if |associatedDomain Set ∩ whitelist |= 0
9:     ret = 0;
10:  else
11:    determine url phishing or not via classifier C
12:    if C (url) = 0
13:      ret = 0;
14:    else
15:      ret = 1;
16:      put domain(url) into blacklist;
17:    end if
18:  end if
19: end if
20: end if
```

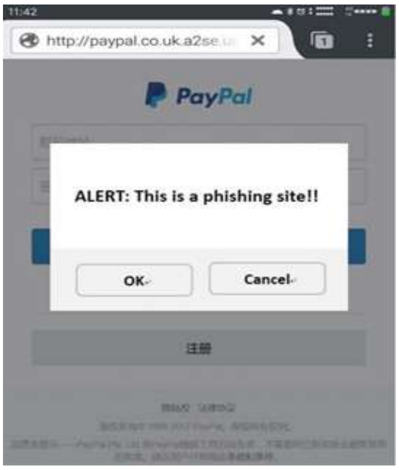
Output:
ret: url phishing (1) or not (0)

Algorithm is particularly best for web browser plugins. The brand resource request relation analyzing steps step 1 - step 9 can be executed in linear time. Only a very small number of URLs to be detected need to execute subsequent steps. BlackPhish can automatically extend the blacklist (step 16), which will be a useful complement to the blacklisting method. C could be different algorithms with different complexity, for different application scenarios, such as heuristic rules or machine learning algorithms.

IV. Algorithm Implementation And Evaluation

Some phishing sites implicitly call the brand resources in CSS and JS files, in this Section. Every resource file needs to be retrieved, so directly matching the brand domain names is complicated. The resource requests analysis becomes easier in the Web browser scene. The external resources request can be listened to when accessing the Web page. Taking into account that more and more phishing URLs are designed for the mobile Web, we used Web Extensions to develop a Firefox extension for Android [5].

To intercept HTTP requests, we used the web Request API, which can listen to various stages of making an HTTP request. The Plug-in works without additional network load considering that the browser has to load all kinds of resources when it accesses a Webpage. Heuristic classifier C is used to speed up the identification of suspected phishing sites. The heuristic rules checks whether the page contains input box, sensitive words, and copyright notice and so on.



ALERT: This is a phishing site!!

Fig. 1 shows an example of an alert message given by our Approach.

V. Conclusion And Future Work

Phishing attacks become even more violent with the popularity of mobile consumption and electronic payment. This paper analyzes the dependence of phishing websites on brand resources, and proposes a novel solution –BlackPhish our novel approach to make a robust phish recognition model. Experiment results showed that BlackPhish can effectively detect phishing sites. It can be used as an upgraded version of the blacklist method. In the future, we can combine Logistic regression technique with our novel approach to make a robust phish recognition model.

Acknowledgements

I would like to express my special thanks of gratitude to my colleagues and also our principal (Dr. Md. Ramzan) who gave me the golden opportunity to do this wonderful project on the topic (Anti-Phishing Detection), which also helped me in doing a lot of Research and i came to know about so many new things I am really thankful to them.Secondly, I would also like to thank my parents and friends who helped me a lot in finalizing this paper within the limited time frame.

References

[1]. FBI Warns of 270% Increase in BEC Scams, $2.3 Billion Lost. http://www.batblue.com/fbi-warns-of-270-increase-in-bec-scams-2-3-billion-lost/.

[2]. Xiang. G, Hong. J, Rose. C. P, &Cranor, L. F. (2011). CANTINA+: A Feature-Rich Machine Learning Framework for Detecting Phishing Web Sites,*ACM Transactions on Information and System Security*, 14(2).

[3]. Sheng. S, Wardman. B, Warner. G, Cranor L. F, Hong. J, & Zhang. C, (2009). An empirical analysis of phishing blacklists. *In Proceedings of Sixth Conference on Email and Anti-Spam (CEAS)*.

[4]. Prakash. P, Kumar. M, Kompella. R. R, & Gupta. M. (2010). Phishnet: predictive blacklisting to detect phishing attacks,*In INFOCOM, 2010 Proceedings IEEE (pp. 1-5). IEEE.*

[5]. Legacy extensions for Firefox for Android - Mozilla | MDN. https://developer.mozilla.org/en-US/Add-ons/Legacy_Firefox_for_Android