© 2018 IJSRSET | Volume 4 | Issue 4 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Phishing Detection Based on CSS Features of Web Pages

Niyati Raj*, Prof Jahnavi Vithalpura
IT Department, L. D. College of Engineering, Ahmedabad, Gujarat, India

ABSTRACT

Since last decades, online technologies have revolutionized the modern computing world. However, as a result, security threats are increasing rapidly. A huge community is using the online services even from chatting to banking is done via online transactions. Customers of web technologies face various security threats and phishing is one of the most important threats that need to be address. Therefore, the security mechanism must be enhanced. The attacker uses phishing attack to get victims credential information like bank account number, passwords or any other information by mimicking a website of an enterprise, and the victim is unaware of phishing website.

Keywords: Phishing Attack, CSS features, Web Security

I. INTRODUCTION

Phishing is an attack in which an attacker creates a fraudulent website which looks exactly like the original one to obtain sensitive information of users like password, user id and card details. The attackers use email tricks to send malicious link or some attachments which including the extraction of username and passwords from victims. Social engineering techniques used for phishing attack to exploits weakness in security of web. Attacker sends malicious emails to trick users to visit their fraudulent websites and to enter their credentials such as user name, passwords or other credit card details. These fraudulent websites created only to collect sensitive information from victim users.

Types of Phishing Detection Techniques:

1.1 Black-list and White-List Based Detection

Black-list and White-List based detection technique is most widely used in web browsers like Mozilla's Firefox and Google Chrome safe browsing API. White list and black list based detection technique is very easy to implement on browsers, that's why it is used widely as extensions or browser toolbars. In spite of this Blacklist and white list based detection it is not sure to get full privacy protection of user's data and returns high false negative rate because of the shortest lifetime of fraudulent web sites. Black-list and white-list based solutions is not effective on that web pages which are previously unseen. So it is needed to update and maintain the white-list and black-list for better accuracy and this is the main weakness of this technique.

1.2 URL based detection

URL based phishing detection technique focus on URLs of webpage and compares each URL of web pages to detect malicious websites. This technique works on URL features such as size of URL, number of dot, number of hyphen, number of characters, numeric and variables. Some attacker keeps same URL as original website's URL and makes some minor changes in it. Using this method we can easily detect websites which have most similar pattern.

1.3 Content based detection

The content-based phishing detection technique extracts all keywords from a webpage, and uses them in an Internet to search and identify the original website, and compares the suspicious site with the original website. It doesn't uses any white-lists or black-lists, eliminates the need to maintain such lists.

1.4 Phishing detection based on other features

There are some other methods available to detect phishing website which works on a basis of a machine-learning approach WOT and iTrustPage [11]. It checks a webpage is a phishing page or not is decided based on page reputation, these are either derived from other web pages or by the community of anti-phishing. Compared to this category, proposed method is based on the features like CSS of web pages. Attacker keeps same property of CSS to maintain its look and feel so visual features are not easily changed.[1]

The organization of this document is as follows. In Section 2 (**Proposed Methods**), we will give detail of all phishing detection methods. In Section 3 (**Results**), presents chrome extension working and detection process. In Section 4(**Conclusion**), there is a conclusion of our method.

II. PROPOSED METHOD

Our proposed approach uses CSS as the basis to accurately quantify the visual similarity of each page element. Most of time attacker directly copies CSS of victim page and pasted it into their fraudulent web page.

CSS Extraction

CSS structure contains Selectors and Selectors have their Property and value, which can be represented as
Selector-1[Property-1: Value-1, Property-2: Value-2]
Selectors can be of type Id, Class, Tag and others.
For example,

```
div { color: #00ff23;}
.class { font-size: 10px;}
#id3 { height: 50px;}
```

Here, we will match the selectors and their property and values to the other page's CSS structure. But sometimes attacker add some bugs in their CSS which have property with none or hidden values. It doesn't make any difference in visual appearance to the web page but when we compare it with each other, it shows negative result. This is the main drawback of CSS features. So in our proposed algorithm, we have eliminated properties which contain any none or hidden values.

The proposed algorithm takes two pages, including page contents and layout specifications, and output the similarity score based on the visual layout similarity.

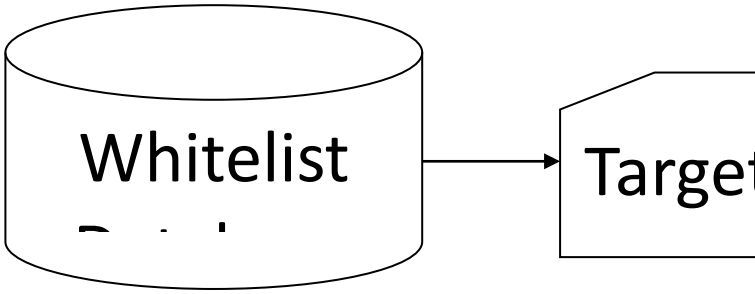

Figure 1. Architecture of Proposed Method

In this method first we have to create white-list database, which contains URLs of websites. In pre-processing, first we extract CSS of white-list web pages and computes vector of it.

We have implemented our proposed method as a Google Chrome Extension, which works in three phases.

Phase 1: Extract Features

In the first phase, whenever user opens any web page then it will check that whether the page contains any input parameters or not. If it doesn't have any input parameter then it will be declared as a non phishing page and then no need to check whether it is phishing or not. And if the page contains any input parameters then it will passes through this phases.

That page will be extracted in its CSS structure and converted into the vector.

Phase 2: Comparison

In the second phase, based on the CSS features, match the similarity between the suspicious page and the pages in the white list database.

Phase 3: Result

Finally, shows the result by comparing the pages' similarity scores to a preset threshold. If the similarity scores are beyond and there exist other clues indicating that two testing pages are different, the suspicious page will be considered as a phishing page and detection message will be displayed on page.

III. RESULTS AND ANALYSIS

Chrome extension and is working is shown in below figures.

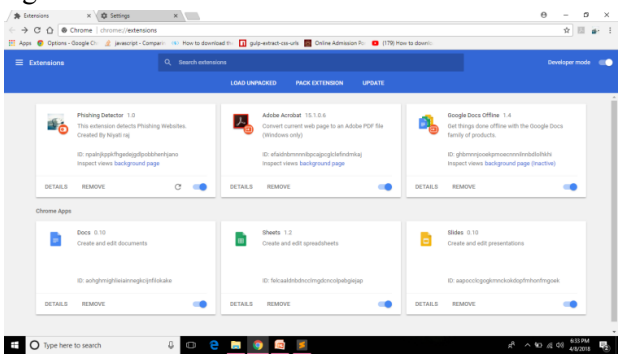

Figure 3.1. Phishing Detector

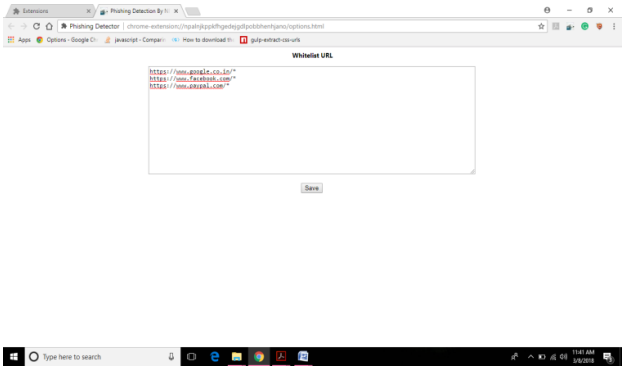

Figure 3.2. White-list Database

Above figure shows white-list database, in which user enters URLs of white-list websites.

Now, we got phishing pages from phishtank.com and then we have checked performance of an extension for 100 web pages.

Below figure shows working of Phishing Detector.

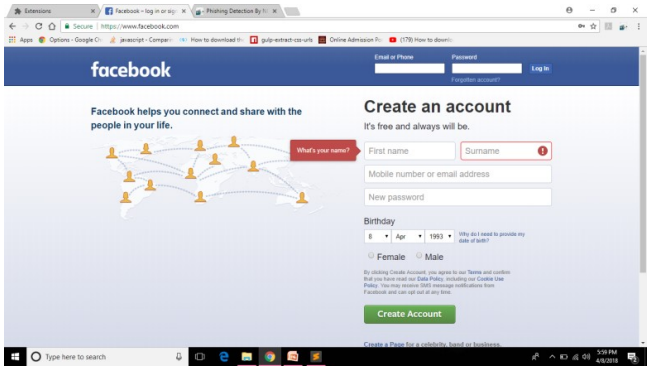

Figure 3.3. Original web page

Figure 3.4. Detected Phishing web page

True Positive Rate and True Negative Rate

From hundreds of phishing pages Phishing detector has detected 91 web pages as a phishing page. That's why true positive rate of Phishing Detector is 91 and True Negative Rate of Phishing Detector is 94.

IV. CONCLUSION

After doing literature survey on different techniques to detect phishing web pages we have concluded that visual similarity based detection technique is the most efficient technique. In proposed method we have used visual similarity of web pages, which is based on CSS features of web pages. Proposed method compares CSS of white-list web page and suspicious web page CSS and then generate similarity score between them. We implemented Phishing detector as an extension of Chrome Browser and checked its effectiveness using Phishing pages.

V. REFERENCES

[1] "Bait Alarm: Detecting Phishing Sites Using Similarity in Fundamental Visual Features" by Jian Mao, Pei Li, Kun Li, Tao Wei, and Zhenkai Liang in 2013 IEEE 5th international conference on intelligent networking and collaborative systems

[2] "Utilisation of website logo for phishing detection" Kang Leng Chiew, Ee Hung Chang, San Nah Sze,Wei King Tiong in 2015 Elsevier

[3] "Visual Similarity based Anti-Phishing with the combination of Local and Global Features" by the Yu Zhou, Yongzheng Zhang , Jun Xiao, Yipeng Wang, Weiyao Lin in 2014 IEEE 13th international Conference on Trust, Security and Privacy in Computing and Communication

[4] "Use of HOG Descriptors in Phishing detection" by ahmet Selmen Bozkir and Ebru akcapinar Sezer in year 2016, 4th international symposium on digital forensics and security (ISDFS'16)

[5] "A computer vision technique to detect Phishing Attacks" by Routhu Srinivasa Rao, Syed Taqi Ali, year 2015, 5th international conference on communication systems and network technologies

[6] "B – APT : Bayesian anti – phishing Toolbar" by Peter Likarish and Eunjin (EJ) Jung, Don Dunbar, Thomas E. Hansen and Juan Pablo Hourcade

[7] Phishing [online] available: https://en.wikipedia.org/wiki/Phishing

[8] Phishing attacks [online] available: https://digitalguardian.com/blog/what-phishing-attack-defining-and-identifying-different-types-phishing-attacks

[9] Phishing attacks [online] available: https://www.pcworld.com/article/135293/article.html

[10] Safe browsing [online] available: https://developers.google.com/safe-browsing/

[11] Anti phishing tool [online] available: https://www.microsoft.com/enus/research/publication/itrustpage-a-user-assisted-anti-phishing-tool

[12] Chrome Extension Guide [online] available: https://developer.chrome.com/extensions