International Journal of Advance in Computer Science & its application – IJCSIA 2018
Copyright © Institute of Research Engineers and Doctors , SEEK Digital Library
Volume 7 : Issue 2 [ISSN : 2250-3765] - Publication Date: 25 June , 2018

Website Forgery/Falsification Detection Technique using Hyperlink Information

[Ji-Ho Cho, Chung-Hyun Lim, Hyo-Jung Ahn, and Geuk Lee*]

***Abstract*— In this paper, we propose a forgery/falsification detection technique of an web site using hyperlink information in the web site. The system crawls all hyperlink information of the web site when a user accesses to the suspicious web site that has the financial information stealing purpose. The captured multiple hyperlink information is compared with those of normal web site hyperlinks information to detect forgery/falsification. The proposed system calculates distance of the normal site hyperlink strings with captured one using Levenshtein distance algorithm to detect whether the site is normal or not. If it is determined as normal, analysis procedure is finished. But if it is determined as abnormal, a warning message is sent to the user to prevent additional financial information spill and further accidents from the forgery/falsified web site.**

***Keywords*— Website Forgery/Falsification Detection, Pharming, Hyperlink.**

I. Introduction

Users of financial services using online, has increased continuously for convenient. In South Korea, based on the end of March 2015, the number of registered internet banking customers (including mobile banking) is 108.61 million people. This is an increase of 5.3% compared to the previous quarter end. The number of daily usage of internet banking (including mobile banking) is average 769.4 million. This is increased by 8.6% compared to the previous quarter [1].

Ji-Ho Cho / Dept. of Computer Engineering
College of Engineering / Hannam University
Republic of Korea
charismaup@nate.com

Chung-Hyun Lim / Dept. of Computer Engineering
College of Engineering / Hannam University
Republic of Korea
Lim147741@gmail.com

Hyo-Jung Ahn / Dept. of Computer Engineering
College of Engineering / Hannam University
Republic of Korea
Hyojung0168@gmail.com

Geuk Lee / Dept. of Computer Engineering
College of Engineering / Hannam University
Republic of Korea
leegeuk@hnu.kr

An increase in the user's Internet banking, made the result of exposure easily to the threat of hackers. Hackers make the same pharming site as the actual internet banking site. Hackers use this pharming site so as to induce a user to enter the personal information that can be used for internet financial transactions. This process is similar to the actual banking transaction process. Therefore, the user has difficulty to find the forgery or falsification website.

The attack using forged/falsified website is expected to appear as a variety of patterns in the future. In this paper we analyzed the type of website forgery attacks. And propose a system that can detect and prevent web forgery/falsification efficiently.

It is assumed that the user's system has been infected with malicious code that has the financial information taken purpose through the falsified or forged website. By comparing the similarity of the captured images of normal website and the images of current website, the system decides whether the website is forged/falsified. Finish the analyzing when the comparing result is determined to be normal. If it is determined to be abnormal, the system sends the warning message to the user.

As a result, through the detection of falsified or forged website, it prevents monetary damages and personal information spill.

II. Related Studies

A. *Phishing and Pharming Attacks*

Phishing is a compound word of Personal data and Fishing. Phishing is a technique that requires financial transaction information available on the web page by phone or email. It sends a call or e-mail disguised as financial institutions or authorized public institutions.

Pharming is a compound word of Phishing and Farming. Pharming should operate your PC infected with malware. Pharming, even if the victim is connected to a normal site will be connected to a fake bank site. Then, after the financial transaction information intercepted gives monetary damages.

An attacker hacked a large number of websites to find an insecure website. The attacker infects Pharming malicious code on this insecure websites. When the security vulnerable user is accessed to this infected website, malicious code is installed to user. Recently, representatively using a floating banner technique attempts to attack trying to impersonate the Financial Supervisory Service. Pharming site contained in the banner is a forged banking site. Through this an attacker aims to steal the users' personal information (certificate, security card number, user ID and password).

International Journal of Advance in Computer Science & its application – IJCSIA 2018
Copyright © Institute of Research Engineers and Doctors , SEEK Digital Library
Volume 7 : Issue 2 [ISSN : 2250-3765] - Publication Date: 25 June , 2018

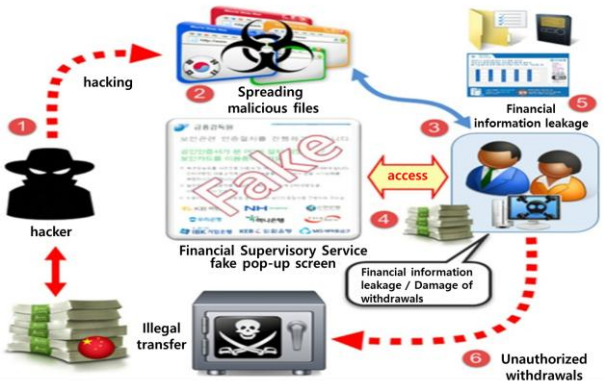

Figure 1. Disseminate a pharming malware by impersonating the Financial Supervisory Service

Figure 3. Website disseminated morphology of pharming malware infections

B. *Hosts File and Hosts.ics Files Tampering*

Hosts file records the IP address corresponding to URL address in the file. Hosts file is check whether the URL in the previous access to the DNS server exists in the list of the hosts file. And it connects to the IP address corresponding to the URL. Hosts.ics file is to specify a function to force the network address for the system at the time of ICS(Internet Connection Sharing). This file is a higher priority than normal hosts file. So if the hosts file exists with hosts.ics file, refer to the hosts.ics file first. If the Hosts.ics file does not exist, refer to the hosts file [2].

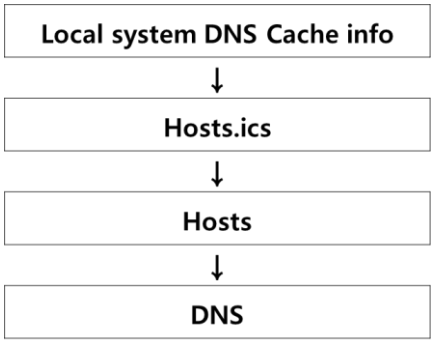

Figure 2. Priorities that are referenced to access the website

C. *Pharming Attacks using the Floating BannerTtechnique*

Floating banner attack is the technique to insert Pharming ad in ads banners. Clicking the banner ads is a malicious code that connects to the pharming sites. The attacker hacks numerous websites to find the insecure website. Attacker infects the vulnerable website with Pharming malware. If the security vulnerable users access the infected website, malware is installed in the user system. Recently, representatively using a floating banner technique attempts to attack trying to impersonate the Financial Supervisory Service. Pharming site contained in the banner is a forged banking site. The aim of this attack is to steal the users' personal information (certificate, security card number, user ID and password) [3]

III. Techniques Used in the System

A. *jsoup*

For the hyperlink data collection, the sysytem uses jsoup, a Java library for HTML work [5]. It crawls the page currently accessed by the user. The system collects hyperlink data in the web page using jsoup.

```
*링크데이터 수집*
https://www.wooribank.com/#introHome
https://www.wooribank.com/#introNav
https://w1spot.wooribank.com/pot/Dream?withyou=CQIBG0050
https://spib.wooribank.com/pib/Dream?withyou=CMLGN0001
https://sbiz.wooribank.com/biz/Dream?withyou=CMLGN0002
https://www.wooribank.com/#none
https://www.wooribank.com/#content
https://www.wooribank.com/#introMall
https://www.wooribank.com/#introFinance
https://spot.wooribank.com/pot/Dream?withyou=PODEP0001
https://spot.wooribank.com/pot/Dream?withyou=ln
https://spot.wooribank.com/pot/Dream?withyou=fx
https://spot.wooribank.com/pot/Dream?withyou=fn
https://spot.wooribank.com/pot/Dream?withyou=is
https://svc.wooribank.com/svc/Dream?withyou=rp
https://www.wooribank.com/#none
https://spib.wooribank.com/pib/Dream?withyou=CMLGN0001
https://sbiz.wooribank.com/biz/Dream?withyou=CMLGN0002
https://www.wooribank.com/#none
https://spib.wooribank.com/pib/Dream?withyou=ct&fromSite=pib
https://sbiz.wooribank.com/biz/Dream?withyou=ct&fromSite=biz
https://spib.wooribank.com/pib/Dream?withyou=ps
https://spib.wooribank.com/pib/Dream?withyou=PSINQ0001
https://spib.wooribank.com/pib/Dream?withyou=PSTRS0001
https://svc.wooribank.com/svc/Dream?withyou=PSTAX0001
https://spib.wooribank.com/pib/Dream?withyou=PSDEP0010
https://spib.wooribank.com/pib/Dream?withyou=PSFND0001
https://spot.wooribank.com/pot/Dream?withyou=is
https://spib.wooribank.com/pib/Dream?withyou=PSLON0001
https://spib.wooribank.com/pib/Dream?withyou=PSFXD0002
https://spib.wooribank.com/pib/Dream?withyou=PSTRT0086
```

Figure 4. Hyperlink data crawling

"Fig. 4" shows the results of hyperlink data collection by crawling the main page of a certain financial site using jsoup. The number of links collected from the financial site used as a sample was larger than 250, and "Fig. 4" shows a part of them.

B. *Levenshtein Distance Algorithm*

The name Levenshtein Distance algorithm was derived from the name of Russian scientist Vladimir Levenshtein. The Levenshtein Distance algorithm is also called edit distance algorithm. The Levenshtein Distance algorithm is an algorithm designed to measure the similarity of two character strings. The Levenshtein Distance algorithm is used for spell checking, speech recognition, and plagiarism detection. The Levenshtein distance algorithm compares two

character strings using a two-dimensional array and conducts insertion, deletion, and changes in each character string to obtain a minimum edit distance value. The accumulated values of minimum edit distances obtained for individual parts using the Levenshtein distance algorithm becomes final edit distance value of the two character strings. The final edit distance value is used as a measure of similarity. [6].

TABLE I. OPERATION PROCESS OF LEVENSHTEIN DISTANCE ALGORITHM

		A	L	L	I	G	A	T	O	R
	0	1	2	3	4	5	6	7	8	9
E	1	1	2	3	4	5	6	7	8	9
L	2	2	1	2	3	4	5	6	7	8
E	3	3	2	2	3	4	5	6	7	8
V	4	4	3	3	3	4	5	6	7	8
A	5	4	4	4	4	4	4	5	6	7
T	6	5	5	5	5	5	5	4	5	6
O	7	6	6	6	6	6	6	5	4	5
R	8	7	7	7	7	7	7	6	5	4

"TABLE I" shows the process of the Levenshtein distance algorithm, which compares ALLIGATOR with ELEVATOR to obtain the edit distance. When ALLIGATOR and ELEVATOR are compared, if the values are equal, the value above the left at the upper part of the left diagonal line should be brought. When ALLIGATOR and ELEVATOR are compared, if the values are different, the smallest value among the values on the top, on the left, and above the left diagonal line plus 1 should be brought. When the operation of the Levenshtein Distance algorithm has been completed, the value at the right bottom becomes the edit distance of the two character strings.

IV. Website Forgery/Falsification Detection Techniques using Hyperlink Information

In this paper, we propose a website forgery/ falsification detection system using hyperlink information. Assume that a user's system is infected with malware, with the purpose to steal your financial information through forged /falsified websites. Users are connected to the forged/falsified website by the malware instead of normal website, the detection system compares link information of the normal website with currently connected website to detect the forgery/falsification. If forgery/falsification is detected, it send warning message to the user and prevents the monetary damages and personal information spill from the site.

A. System Configuration and Operation

The proposed system consists of four parts. Crawling module, similarity check module, result analysis module and message sending module "Fig. 5".

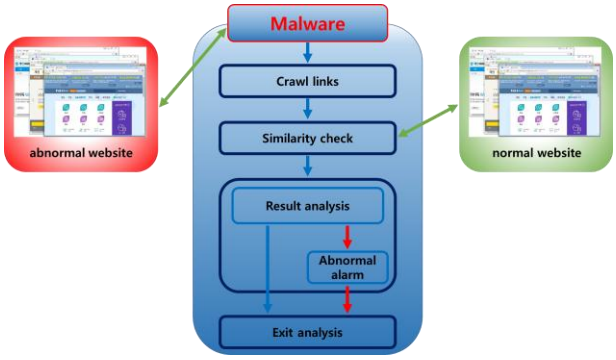

Figure 5. Website forgery/falsification detection system

If a host system is infected with malware from hacker access to financial websites, the crawling module of the system collects hyperlink data. This web page hyperlink information is compared with the normal financial web page information. On the basis of the result of the comparison it is determined whether to send a warning alarm to the user or finish the analysis. Hyperlink information of normal financial web page is crawled in advance and stored in the system. Financial Websites that are analyzed when the user accessed is most of all online banking websites of South Korea. System manager can changes and designates these target websites to be analyzed.

If a user connects to a certain financial website, hyperlink information of the web page crawled by the system for similarity comparison. Each time a user connects to certain financial website, the main page, login page, security card number entering page are crawled. And it also crawls the web page whenever user moves to another page. The hyperlink crawl module uses jsoup. Hyperlink information of normal website was also saved in advance using jsoup.

In similarity check module, crawled hyperlink information is compared with normal hyperlink information which was saved in advance. Comparison uses Levenshtein distance algorithm. If average distance is within the boundary, the result analysis module outputs a "1"and this means normal. If average distance is out of the boundary, the result analysis module outputs a "0"and this means abnormal. When the result is abnormal, it notifies to user. The result is "1", analysis process is finished. The result is "0", it notifies to user using 'abnormal alarm' and prevents information spill from malware.

V. Experiment Result

A. Experiment Environment

Real internet banking websites and used by the number of actual users and real forgery site are used for experiment. We use WINDOWS 10 operating system and use jsoup library for hyperlink data capture. And we use JAVA as the system developing language.

B. Hyperlink Information Capture

jsoup is used to collect hyperlink information in a page. Captured information is main page of internet banking site, login page and money transfer page. The system compares newly crawled hyperlink information with saves normal one for the analysis. Sites used for experiment are shown in "Fig. 6".

Figure 6. Comparison of website *(left ones are saved normal website, right ones are newly captured site)*

C. Similarity Check

"Fig. 7" is the results of the similarity check process applying Levenshtein distance algorithm.

```
http://www.woorifis.com/                    http://www.woorifis.com/
http://www.wooricredit.co.kr/               http://www.wooricredit.co.kr/
http://www.woorifoundation.or.kr/           http://www.woorifoundation.or.kr/
http://www.woorifs.co.kr/                   http://www.woorifs.co.kr/
http://www.wooriмiso.or.kr/                 http://www.wooriмiso.or.kr/
http://www.wfri.re.kr/                      http://www.wfri.re.kr/
http://www.wooriib.com/                     http://www.wooriib.com/
https://www.facebook.com/wooribank          https://www.facebook.com/wooribank
https://www.twitter.com/wooribank           https://www.twitter.com/wooribank
https://www.wooribank.com#none              https://www.wooribank.com#none
https://www.wooribank.com#none              https://www.wooribank.com#none
https://www.wooribank.com#none              https://www.wooribank.com#none
https://www.wooribank.com#none              https://www.wooribank.com#none
https://www.wooribank.com#none              https://www.wooribank.com#none

*링크데이터 분석결과*                       *링크데이터 분석결과*
1.0                                         0.8701803051317615
링크데이터가 동일합니다.                    링크데이터가 다릅니다.
```

Figure 7. Result of similarity comparison

The result of comparing three sites. According to the result, the similarity of the site 1 and site 2 may are not differences. However, site 3 is analyzed to be different as a result of comparison to two link information. Site 3 page is concerned with security number entering page. Newly crawled page which is connected currently by user is forgery website page. And the page requires inputs of a security card number. The link information of this web page is slightly different from to the normal websites. In addition, this page induces to enter the full set of numbers in security card.

D. Result Analysis

"TABLE II" shows a hyperlink information similarity comparison and analysis result. Site 1 and site 2 is analyzed as normal but the site 3 is analyzed as abnormal. As shown in "Fig. 8", if the result is determined as abnormal, system sends waring alarm message to the user.

TABLE II. RESULT OF ANALYSIS

	normal	abnormal	result	Process according to the result
site 1	1	0	normal	Finish the analysis process
site 2	1	0	normal	
site 3	0	1	abnormal	Send warning alarm to the user

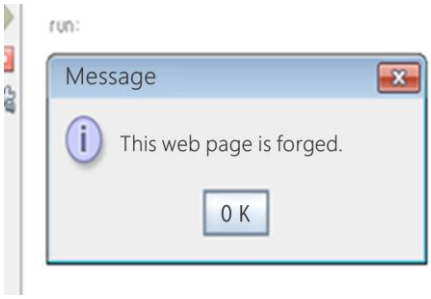

Figure 8. Warning Message to the user

VI. Conclusion

On Users of financial services using online have increased continuously due to the convenient. The increase of the online financial service usage, like Internet banking, made the result of exposure easily to the threat of hackers.

The website the forgery/falsification detection method using hyperlink information was proposed in order to detect whether or not forgery/falsification of the website. The proposed system crawls link information of the website when a user accesses to the suspicious forgery/falsification website which has the financial information stealing purpose. The crawled hyperlink information is compared with that of normal website information to detect forgery/falsification. The proposed system calculates similarity value of normal site hyperlink information with the captured one to detect whether the site is normal or not. If it is determined as normal, analysis procedure is finished. But if it is determined as abnormal, a warning message is sent to the user to prevent additional financial information spill and further accidents from the forgery website.

Further study is needed to increase the speed of comparison process and the accuracy of detection. This system can be used as an assistant tool in cyber security monitoring system to make security monitoring person response quickly when a forgery/falsification occur.

Acknowledgment

This research was supported by Human resources Exchange program in Scientific technology through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and future Planning (NRF - 2016H1D2A2916091).

*Geuk Lee is corresponding author.

References

[1] Jung-Hyuk Kim, "The present situation of use of domestic Internet banking services in the third quarter of 2015", Bank of Korea press release, November 2015.

[2] http://www.hauri.co.kr/information/issue_view.html?intSeq=243&page=1, Pharming technology appeared in 2014.

[3] http://blog.alyac.co.kr/285, Pharming technique that becomes much tricky to impersonate the Financial Supervisory Service.

[4] Kyu-il Kim, Sang-soo Choi, Hark-soo Park, Sang-jun Ko and Jung-suk Song, "Website Falsification Detection System Based on Image and Code Analysis for Enhanced Security Monitoring and Response", Journal of the Korea Institute of Information Security and Cryptology, vol. 24, no. 5, p. 871-883, 2014.

[5] Liu Quanzhi and Yu Zjilou, "Design and Implementation of Web Information Extraction Sysytem Based Heritrix and Jsoup", Journal of Shandong Normal University, vol.30, no.2, p.16-19, 2015.

[6] Black, Paul E., ed., "Levenshtein Distance", Dictionary of Algorithm and Data Structure[online], 2008.