AN ANTI-PHISHING TOOL TO VERIFY URLS IN EMAIL'S CONTENT

Melad Mohamed Al-Daeef, Nurlida Basir and Madihah Mohd Saudi
Fakulti Sains dan Teknologi, Universiti Sains Islam Malaysia (USIM), Nilai, Negeri Sembilan
E-Mail: meladmohalda@gmail.com

ABSTRACT

Phishing is a threat that causes Internet users to lose the control over their accounts. A variety of anti-phishing approaches have been proposed, one of this variety is anti-phishing client-side tools. Most of these tools are rely on black/white list and heuristic methods. Most of these tools however, still unable to catch all phishing emails, especially zero-day attacks. To success, phishers usually take advantage of weaknesses in implemented anti-phishing methods and human vulnerabilities. Therefore, there is an urgent need for a solution that integrates anti-phishing technical and non-technical (user awareness) approaches. The proposed system (tool) in this paper is an attempt to achieve that goal. The proposed tool is designed to work at user's email browser/platform, and provide the user with basic information about any URL in email's content before clicking it, and probably browse suspicious site. WHOIS query method is employed by the proposed system to obtain required information about URLs in email's content.

Keywords: phishing, user awareness, client-side tool, pre-browsing protection, WHOIS query.

INTRODUCTION

Phishing is a form of cyber crime in which phishers trying to illegally get control over users' accounts by persuading them to visit predesigned fake websites. Phishers usually employ emails to deliver fake URLs (links) to targeted victims. Once clicked, the fake URL will take the user to phisher's website which looks exactly similar to the original website (Watson et al., 2005). Anti-Phishing Working Group (APWG) has reported that, millions of URLs were used for phishing attacks in the second half of 2013. There are many reasons for the success of phishing attacks. First, there still some limitations associated with most proposed anti-phishing technical methods. Phishers usually take advantage of such limitations when performing phishing attacks. The second reason is the lack of Internet users' awareness and knowledge about online threats. Phishers take advantage of human vulnerabilities to overcome anti-phishing technical obstacles (Wilson et al., 2011).

Due to the lack of awareness about online threats, many users are just arbitrarily click obfuscated URLs in emails' content. Thus, it would be better to find an anti-phishing solution that integrates technical and non-technical approaches and protect users from phishing attacks before they reach phishing sites (i.e. provide users with a pre-browsing level of protection).

There are many features that have been used to identify phishing emails. In many cases however, these features were improperly chosen, thus they produce high misclassification results (Toolan and Carthy, 2010). Melad et al., 2014 have argued that, for the feature to be relied-upon in the process of detecting phishing emails, it is better to first evaluate its efficiency in that task. Researchers in the same study have concluded that, employing URLs as a feature can help in producing an accurate decision about questioned email.

A pre-browsing level of protection can be achieved by enabling users from checking the authenticity of URLs embedded in phishing emails before these URLs being clicked. The proposed system in this paper is a client-side tool designed to work at user's email browser/platform that provide the user with important information about URLs in emails' content by utilizing WHOIS query method. WHOIS is a query/response method that is widely used to find information about networks, domains and hosts (Daigle, 2004). The information about URL obtained from WHOIS query include but not limited to, registrar, registration date, expiry date, owner institution, and origin country of the URL. Such information can assist the user to make a true decision about any URL before clicking it. This approach can help in protecting users against zero-day attacks. Zero-day attack is the attack that performed by the phisher using hosts that not appear in blacklists and were not trained on the old data sample (Khonji, 2012).

BACKGROUND ON ANTI-PHISHING SOLUTIONS

Many anti-phishing approaches have been proposed. These approaches are generally classified according to wherein the attack is occur. They are categorized as, network level protection, authentication, server-side filters/classifiers, prevent against duplication, client-side tools, and user education approaches (Ramanathan and Wechsler (2012).

Client-side tools are designed to work at user’s browser and utilize a variety of methods to identify a web page as either phishing or legitimate. These methods include blacklists (lists of known fraudulent sites), whitelists (lists of known safe sites), heuristics, and community ratings. Some of client-side tools are utilizing combinations of these methods. Most of these tools warn users with a dialog box about suspicious websites. Some examples of these tools are:

CallingID is an anti-phishing toolbar that uses 54 different verification tests in order to determine websites' legitimacy. The tool has an indicator that changes from green for trusted sites, yellow for low risk rate sites, and red for high risk rate and thus probably phishing sites. For the site to be rated, some heuristics are employed, these include an examine of the site's country of origin, length of registration, popularity, user reports, and blacklist data.

NetCraft is an anti-phishing toolbar that uses several methods to determine the legitimacy of a website. It traps suspicious URLs containing ambiguous characters. The toolbar enforces the display of browser navigation controls such as the address bar in all windows to defend against pop-up windows which attempt to hide the navigational controls, it also display site's hosting location. The tool also uses a blacklist method. If the user attempts to access a site that is on the blacklist, a pop-up warning recommends cancelling the access, and display a risk rating for visited site.

Spoof guard is an anti-phishing toolbar that employ a series of heuristics to identify phishing pages instead of using black/white list techniques. This tool checks the current domain name and compares it with a list of sites that have visited by the user. The URL of visited page is also analyzed to detect obfuscation and non-standard port numbers. Page's content is also analyzed, noting if there any password fields with no secure connection, embedded links, and images. Links in the web page itself also analyzed using some heuristics. The toolbar displays a red, yellow, or green icon to warn users about visited websites.

eBay toolbar is a browser plug-in that eBay offers to its customers. It uses a combination of heuristics and blacklists. It has "Account Guard" feature that monitors the domain names of visited sites and provides a warning in the form of colored icon which turns to green if visited site was operated by eBay or PayPal, turns to red if visited site was known as phishing site, and turns to gray if visited site was not operated by eBay or PayPal. Users are allowed to report suspected sites to eBay, reported sites will be verified before they being blocked. Known phishing sites are blocked and a pop-up appears, giving users the option to override the block. eBay toolbar requires no effort on the user's part other than to notice toolbar color changes.

Spoof stick is a toolbar that can be added to both IE and Firefox browsers. It provides basic domain information by displaying website's real domain name to the user. This is useful when spoofed links contain multiple sub-domains. The attacker might use a legitimate looking domain name as a sub-domain to craft spoofed link to lure its victims. For example, if the link *http://patrickbond.co.uk/w/www.chase.com/* is used to trick the user, Spoof Stick displays *patrickbond.co.uk*, so the user notices the actual hosting domain.

IE phishing filter, Internet Explorer users have the option to enable the phishing filter as it is not enabled by default. This built-in phishing filter has a downloaded list of known safe sites, and it does real time checking for phishing sites by verifying URLs with an anti-phishing verification server which hosted by Microsoft. IE phishing filter relies on a blacklist technique, and also uses some heuristics when it encounters a site that is not on the blacklist. If a suspected phishing website was encountered, the user is redirected to a built-in warning message and asked if he/she like to continue visiting the site or to close it. Users also provided with an option of reporting suspected phishing sites, or report that a site has incorrectly been added to the blacklist.

Summary of client-side tools

Generally, most of client-side toolbars are based on black and/or white list methods. The main flaw point associated with blacklist method is the required time for new (zero-day) phishing websites to be reported and hence added to the blacklist. During that time, the attack could take a place. The blacklist method could also lead to a False Negative (FN) results; FN means that, the email/website is incorrectly identified as phishing. The white list method on the other hand is a collection of trustworthy URLs. This method however is a time-consuming process. In addition, this method could lead to a high rate of False Positive (FP) results, thus allowing phishing emails/websites to go through; FP means that, the email/website is incorrectly identified as a legitimate. Besides of the flaw points of the black and white list methods, most of users do not pay attention to warnings displayed by toolbars (Ramanathan and Wechsler 2012), (Dhamija et al., 2005), (Wu et al., 2006).

Some of other client-side tools are heuristic-based techniques, in addition to the consumed time by this method, users in most cases are required to adjust many thresholds of implemented tool. Other client-side tools are specific website tools. eBay anti-phishing toolbar for example, can flag only spoof sites which known by eBay and PayPal.

This section described some of client-side tools that designed to detect phishing websites after the user visit the phishing page, and hence became susceptible to fall prey for phishing attack and went into the phase of giving out sensitive information at false page. The proposed system in this paper is an attempt to prevent users from getting into this phase of risk by enabling them to check the legitimacy of URLs in email's content before these URLs being clicked. In other words, protecting users from browsing suspicious pages (i.e. pre-browsing level of protection).

.

THE PROPOSED SYSTEM

The proposed system (tool) is designed to work at users' email browser/platform. Outlook 2007 is used as a platform for the proposed system since Outlook application can be used as an email manager. Figure-1 shows the pseudo code of the proposed system, whereas Figure-2 shows the flow chart of its operations.

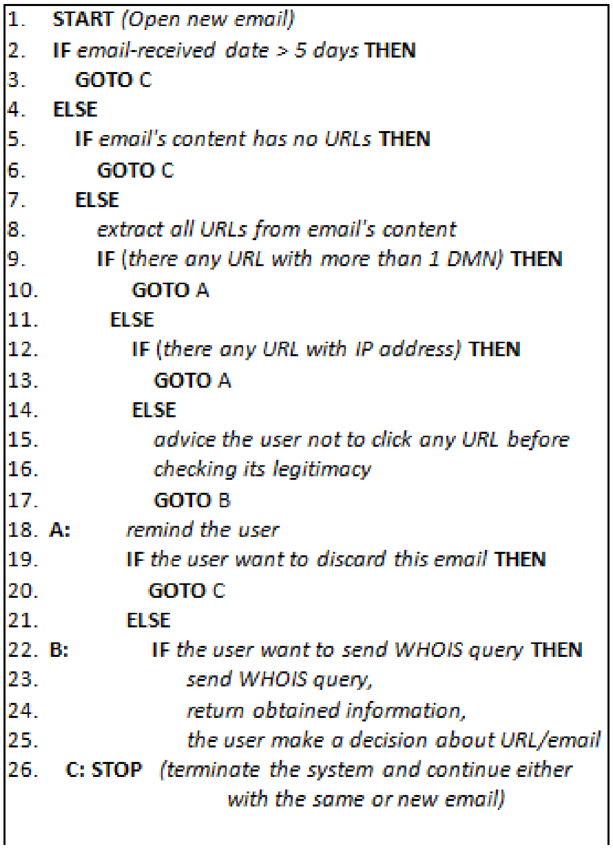
Step No. 1 in the pseudo code shows that, if the email has been received before 5 days, then there is no need of making WHOIS query about any URL in its content, that is due to the fact of that, the average lifetime (threat time) of phishing URLs is about 4 to 5 days (APWG)

Step No. 9 in the pseudo code says that, if there any URL in email's content with more than one domain

name, the user then must be reminded about that suspicious link. Such a link looks like
<ahref =“http://www.profundnet.org/checksessioninfo.php”>https://Genuine.secureregion.com/EBanking/logon/</a>

which appears to be linked to *Genuine.secureregion.com*, the portal of a bank, but it actually is linked to a phishing site *www.profundnet.org*. (Chandrasekaran et al. 2006), (Fette et al. 2007), (Suriya et al., 2009).

Step No. 12 in the pseudo code says that, if there any URL in email's content with an IP address, the user must be reminded about that suspicious link. Any link is said to be suspect when it found to be an IP address rather than a name. For example don’t trust a link if it has delivered to you by email as:
http://212.33.67.194/.citibank/accountexpirycheck.net
(Chandrasekaran et al. 2006), (Suriya et al., 2009).



```
1.  START (Open new email)
2.  IF email-received date > 5 days THEN
3.     GOTO C
4.  ELSE
5.     IF email's content has no URLs THEN
6.        GOTO C
7.     ELSE
8.        extract all URLs from email's content
9.        IF (there any URL with more than 1 DMN) THEN
10.            GOTO A
11.         ELSE
12.            IF (there any URL with IP address) THEN
13.               GOTO A
14.            ELSE
15.               advice the user not to click any URL before
16.               checking its legitimacy
17.               GOTO B
18. A:          remind the user
19.             IF the user want to discard this email THEN
20.                GOTO C
21.             ELSE
22. B:             IF the user want to send WHOIS query THEN
23.                    send WHOIS query,
24.                    return obtained information,
25.                    the user make a decision about URL/email
26.   C: STOP  (terminate the system and continue either
                         with the same or new email)
```

Figure-1. Pseudo code of the main operations of the proposed system.

THE IMPORTANCE OF THE PROPOSED SYSTEM

The importance of the proposed system lies in two main points of advantages. First, the system will assist users in detecting phishing attacks (including zero-day ones) and warns them about the potential risk before they visit phishing website and go into the stage of typing-in sensitive information (i.e. pre-browsing level of protection). The second point of advantage is that, the proposed system is designed to be a user awareness-oriented tool which directly engage users in the process of verifying URLs' legitimacy, as a consequence, overall users' awareness about online threats is increased.

OPERATION EXAMPLE OF THE PROPOSED SYSTEM

In this section, an example of working operation of the proposed system is given. Figure-3 shows a phishing email that asking the user to click a link in email's content. This email pretend came from RHB Bank, this email is asking the user to click the link which appears to the user as *https://logon.rhb.com.my*. A snapshot of the proposed system in Figure-4 shows the extracted URL from the content of this phishing email. Figure-4 however shows that, this extracted URL is different from what was shown to the user, the extracted URL is,
http://sma-yapan.sch.id/okuphp

When a WHOIS query about extracted link was sent by the proposed system, there was no useful information were obtained although many WHOIS servers were queried. However, when a WHOIS query about RHB.COM.MY domain was sent, the information in Figure-5 and other information were retrieved. This information can assist the user to make the correct decision about such an email, or at least the user will be more cautious if he/she has decided to click any URL in the email.

The proposed system provides the user with a function to write the domain name that he/she has a relationship with, and make a WHOIS query about this domain.

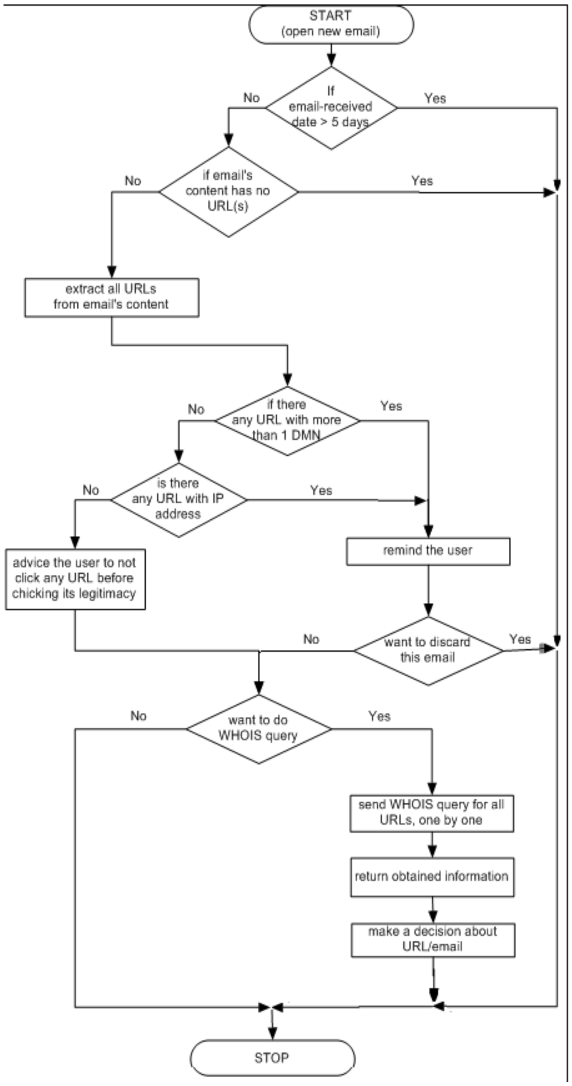
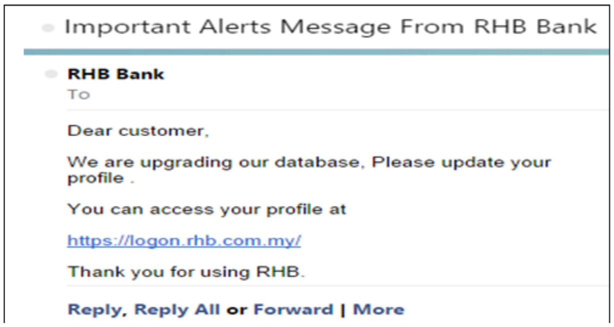

START
(open new email)
If email-received date > 5 days
No
Yes
if email's content has no URL(s)
No
Yes
extract all URLs from email's content
if there any URL with more than 1 DMN
No
Yes
is there any URL with IP address
No
Yes
advice the user to not click any URL before chicking its legitimacy
remind the user
want to discard this email
No
Yes
want to do WHOIS query
No
Yes
send WHOIS query for all URLs, one by one
return obtained information
make a decision about URL/email
STOP

Figure-2. Flow chart of main operations of the proposed system.


Important Alerts Message From RHB Bank

RHB Bank
To

Dear customer,

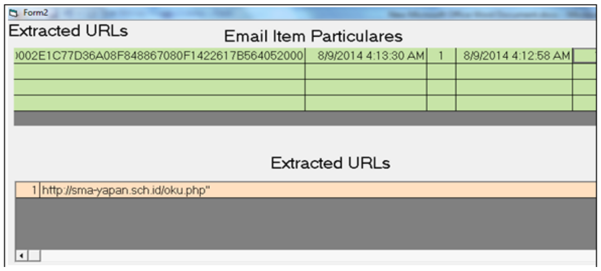
We are upgrading our database, Please update your profile .

You can access your profile at

https://logon.rhb.com.my/

Thank you for using RHB.

Figure-3. Phishing email sample.



Extracted URLs

Email Item Particulares

)002E1C77D36A08F848867080F1422617B564052000	8/9/2014 4:13:30 AM	1	8/9/2014 4:12:58 AM	

Extracted URLs

1	http://sma-yapan.sch.id/oku.php"

Figure-4. Extracted URLs from email's content.

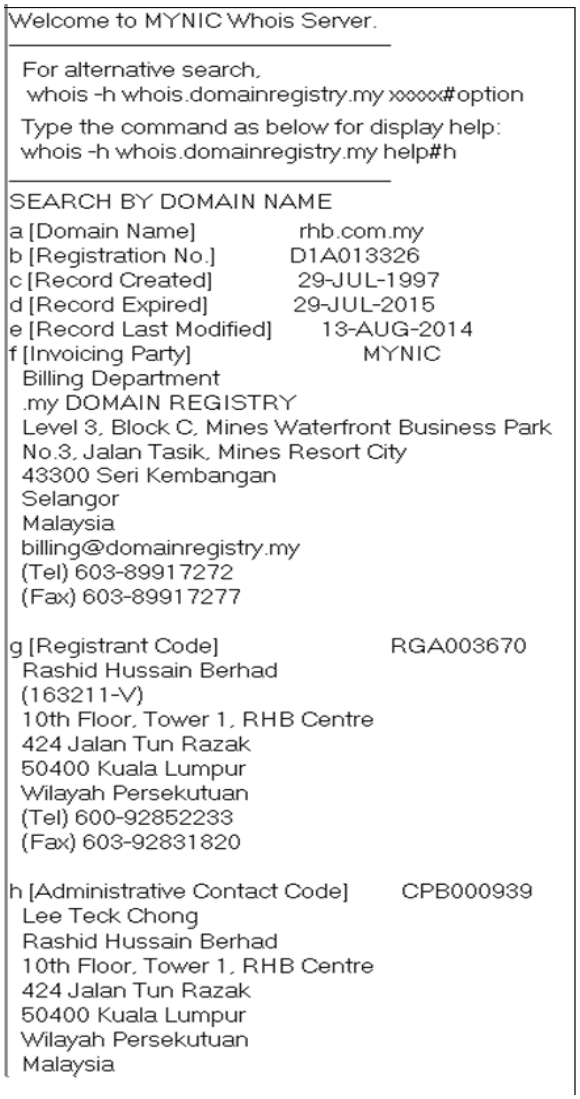
Here we mention that, the information in Figure-4 are not displayed to the user unless the user want to know more details about the questioned URL in opened email. That is to reduce the overhead implementation of the system. Obtained information in Figure-5 are to be summarized in the next development stage of the proposed system to provide the user with a short form of obtained information.

CONCLUSION, FUTURE WORK, AND LIMITATIONS

Unlike other client-side tools that rely on a verity of methods such as black/white lists, heuristics, and etc, the proposed system in its main operation is relying on a live WHOIS query method to provide users with a pre-browsing level of protection including zero-day phishing attacks which still making a serious challenge to most of known anti-phishing client-side tools. End users are targeted to be directly interacting with the proposed system, as a consequence, users' awareness about online threats will be eventually increased.

As a future work, retrieved information about URLs can be summarized and provided to the user in a short form. The system is also can be equipped with a function to report detected phishing URLs. Detected suspicious URLs can be also stored in a local or online user-owned database to remind the user about any URL without making another WHOIS query if same URL was found in other emails.

Although of its promising advantages in protecting users and increasing their awareness about online threats, there still some level of implementation overhead associated with the proposed system. This level of implementation overhead however, still acceptable when compared with its advantages. The implementation overhead is still the nature characteristic of most, if not all other anti-phishing client-side tools, and it is the tax that users must pay for getting protected.



```
Welcome to MYNIC Whois Server.
______________________________________
 For alternative search,
  whois -h whois.domainregistry.my xxxxx#option
 Type the command as below for display help:
 whois -h whois.domainregistry.my help#h
______________________________________
SEARCH BY DOMAIN NAME
a [Domain Name]             rhb.com.my
b [Registration No.]        D1A013326
c [Record Created]          29-JUL-1997
d [Record Expired]          29-JUL-2015
e [Record Last Modified]       13-AUG-2014
f [Invoicing Party]                 MYNIC
 Billing Department
 .my DOMAIN REGISTRY
 Level 3, Block C, Mines Waterfront Business Park
 No.3, Jalan Tasik, Mines Resort City
 43300 Seri Kembangan
 Selangor
 Malaysia
 billing@domainregistry.my
 (Tel) 603-89917272
 (Fax) 603-89917277

g [Registrant Code]                 RGA003670
 Rashid Hussain Berhad
 (163211-V)
 10th Floor, Tower 1, RHB Centre
 424 Jalan Tun Razak
 50400 Kuala Lumpur
 Wilayah Persekutuan
 (Tel) 600-92852233
 (Fax) 603-92831820

h [Administrative Contact Code]      CPB000939
 Lee Teck Chong
 Rashid Hussain Berhad
 10th Floor, Tower 1, RHB Centre
 424 Jalan Tun Razak
 50400 Kuala Lumpur
 Wilayah Persekutuan
 Malaysia
```

Figure-5. Sample of obtained information from WHOIS query.

REFERENCES

Anti-Phishing work Group, http://www.apwg.org/ CallingID–Your Protection from Identity Theft, Fraud, Scams and Malware. http://www.callingid.com/Default.aspx.

Chandrasekaran, M., Narayanan, K., & Upadhyaya, S. (2006, June). Phishing email detection based on structural properties. In NYS Cyber Security Conference (pp. 1-7).

Daigle, L. (2004). WHOIS protocol specification.

Dhamija, R., & Tygar, J. D. (2005, July). The battle against phishing: Dynamic security skins. In Proceedings of the 2005 symposium on Usable privacy and security (pp.77-88). ACM. eBay Toolbar.http://download.cnet.com/eBay-Toolbar/3000-12512_410153544.html?tag=c-ontentMain;downloadLinks.

Fette, I., Sadeh, N., & Tomasic, A. (2007, May). Learning to detect phishing emails. In Proceedings of the 16th international conference on World Wide Web(pp. 649-656). ACM. IE Phishing Filter.http://support.microsoft.com/kb/930168.

Khonji, M., Iraqi, Y., & Jones, A. (2012). Enhancing Phishing E-Mail Classifiers: A Lexical URL Analysis Approach. International Journal for Information Security Research (IJISR), 2(1/2).

Melad Al-Daeef, M. M., Basir, N., & Saudi, M. M. (2014, May). A Method to Measure the Efficiency of Phishing Emails Detection Features. In Information Science and Applications (ICISA), 2014 International Conference on (pp. 1-5). IEEE.Netcraft Anti-Phishing Toolbar.http://toolbar.netcraft.com/.

Ramanathan, V., & Wechsler, H. (2012). phishGILLNET—phishing detection methodology using probabilistic latent semantic analysis, AdaBoost, and co-training. EURASIP Journal on Information Security, 2012(1), 1-22. SpoofStick. [Online]. Available: http://www.spoofstick.comSpoofGuard.http://crypto.stanford.edu/SpoofGuard/.

Suriya, R., Saravanan, K., & Thangavelu, A. (2009, October). An integrated approach to detect phishing mail attacks: a case study. In Proceedings of the 2nd international conference on Security of information and networks (pp. 193-199). ACM.

Toolan, F., & Carthy, J. (2010, October). Feature selection for Spam and Phishing detection. In eCrime Researchers Summit (eCrime), 2010 (pp. 1-12). IEEE.

Watson, D., Holz, T., and Mueller, S. (2005) Know your enemy: Phishing, behind the scenes of Phishing attacks, The Honeynet Project & Research Alliance.

Wilson, C., & Argles, D. (2011, June). The fight against phishing: Technology, the end user and legislation. In Information Society (i-Society), 2011 International Conference on (pp. 501-504). IEEE.

Wu, M., Miller, R. C., & Garfinkel, S. L. (2006, April). Do security toolbars actually prevent phishing attacks?. In Proceedings of the SIGCHI conference on Human Factors in computing systems (pp. 601-610). ACM.