Survey of Anti-phishing Tools with Detection Capabilities

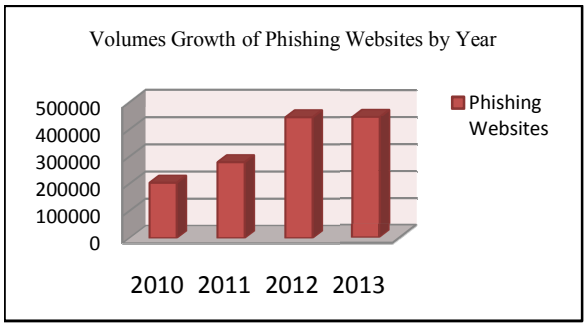
Hiba Zuhair Zeydan[1], Ali Selamat[2], Mazleena Salleh[3]
[1,2,3] Faculty of Computing
Universiti Teknologi Malaysia (UTM), Johor, Malaysia
[1]zzhiba2@live.utm.my, [2]aselamat@utm.my, [3]mazleena@fsksm.utm.my

***Abstract*— Phishers have been continually changed their tricks and emerged novel variants for more security violations and causes of monetary losses in business organizations. The lack of existing anti-phishing solutions considered as an optimum anti-phishing is because of detection incapability specifically against novel phishes. This paper classifies the existing anti-phishing tools, identifies their detection incapability against several kinds of novel phishes and underscores the issues behind this problem. Further it suggests next wave of research to solve it. Targeting academic and industry researchers, this paper provide a valuable source of information to contribute the cyberspace with new products and fulfill the security flaws.**

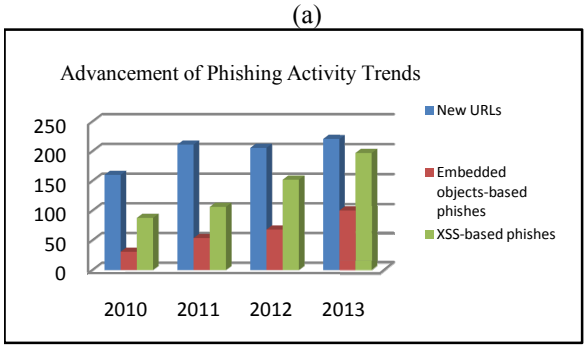
Keywords- Internet phishing; anti-phishing; detection capability; novel phishes.

I. INTRODUCTION

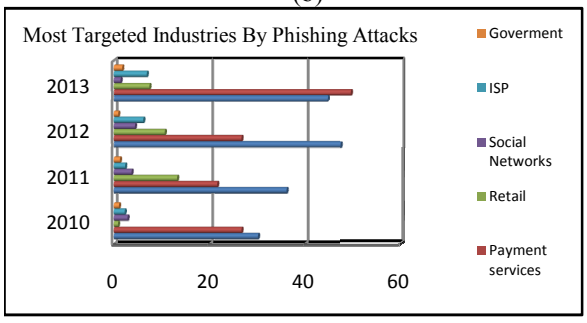
Recently, Internet phishing is one of the most profitable cyber-crime in cyberspace. Internet phishing is exploiting web applications' vulnerabilities and social engineering; phishers disguise the reliable and sensitive transactions of users' and identity theft by impersonating the legitimate websites or delivering phishing emails [1-4]. For its mitigation, various anti-phishing tools have been proposed in the last few years from academic and industry researchers' side. And they have been designed in the form of email filters, anti-virus software, or web browsers plug-in's, add-on's, extensions and toolbars, or as independent web application. However, these tools are in risk which will have consequences in both cyber-security and economy in the future due to the dramatically increase of novel phishes. Novel phishes aim to bypass the existing anti-phishing tools and cause more potential risks such as password harvesting, malware distribution and then more substantial monetary losses [5-8]. As reported by Anti-Phishing Work Group (APWG) which is an international non-profit organization formed in 2003 to keep track current and future phishers' activities, Fig. 1(a) and Fig.1(b) illustrate the rapid growth of phishes in terms of volume and activity year over year since 2010 to 2013 [9]. Consequently, these phishes target many industries in the world such as online payment services, financial organizations, e- banks, retail and ISP services, social networks and online governmental organizations respectively and cause dramatically increased financial losses [1, 2, 10-13], as presented in Fig. 1(c).



(a)



(b)



(c)

Fig.1. (a) Volume growth of phishing attacks, (b) advancement of phishing activities and (c) the most targeted industries by phishing attacks.

978-1-4799-6444-4/14/$31.00 ©2014 IEEE

For the aforesaid issues, none of the notably known phishing tools provides an optimum solution against novel phishes [4, 14-23]. Targeting academic and industry researchers, this survey aims to hang the issues behind the problem of novel phishes and detection incapability of anti-phishing tools against them by characterizing the elements of the problem and presenting the potential areas for further study. The rest of this survey is organized as follows: Section II presents an overview of typical anti-phishing tool in terms of their detection scenarios and application levels. Then, a review of these tools due to their detection capabilities against novel phishes as well as future development of the research is presented in section III. At last, a conclusion is drawn in Section IV.

II. ANTI-PHISHING TOOLS

In the literature most of the existing anti-phishing tools relied on various detection approaches that have been categorized into: lists, heuristics, hybrid, and information flow approaches [1, 7, 8, 15, 21, 24-27].The lists-based approaches include blacklists and whitelists approaches, which rely on frequently updated lists of well-known phishing URLs and legitimate URLs respectively. Whereas, heuristics based approaches predict websites' phishness according to a set of heuristics in website's URL and content. Hybrid approaches have mainly combined the former approaches with the aid of hybrid feature sets and classifiers to detect phishes. On the other hand, information-flow based approaches rely on appending some random credentials before and after the users submit their credentials to a phishing website [1, 2, 7, 10, 15, 18].

Consequently, anti-phishing tools have been implemented at different application levels like: client-side level, server-side level and client-server level [1, 28, 29]. Due to the direct interaction of Internet users with websites via web browsers, they are potentially on risk by phishes. Thus, almost anti-phishing tools are integrated with the popular web browsers such as Google Chrome, Internet Explorer, Mozilla Firefox, Safari and Opera. These integrated tools can keep track of users' activities during web browsing and notify them against any phishing website on real time. However, the web browsers-based anti-phishing tools have some limitations related to the design of intuitive interface, detection accuracy, correct warnings and suitable help system [20, 30].

Whereas, almost phishing email filtering tools are implemented at the server-side. But they are still not effective against web banner advertising, instant chats and messengers which can be exploited by novel phishes. Also some of them rely on visual indicators and fail when users rarely notice the absence and presence of these indicators [28-30]. On the other hand, the anti-phishing tools at client-server structured applications are widely used by commercial organizations like Google, Microsoft and Netcraft but they frequently request for update and maintenance from their databases server [5]. Appendix Table I presents typical anti-phishing tools in terms of some relative merits like related work, year, name, type of approach, type of solution, type of contribution, applied platform and the application level.

As depicted in Appendix Table I, B-APT was developed as white list-based anti-phishing toolbar for US financial institutions and it identified phishing websites on the basis of document object model DOM tokens and Bayesian filter [6]. An automated individual white list-tool AIWL was proposed to protect users' and their online credentials [7]. Likely, some researchers at Google Inc. [31], proposed upgraded Google's phishing blacklist with a classifier as Google Toolbar to identify phishing webpages due to some distinctive features. An enhanced blacklist PhishNet generates new URLs using heuristics and DNS lookup [32]. And, PhishCatch is mainly relied on some weighted rules to classify phishing emails [25]. Then, PhishShark is used to detect phishing websites with the aid of twenty heuristics [34]. Whilst, CANTINA+ was proposed as an upgraded version of CANTINA with the use of one new feature, ten additional features, four features from typical CANTINA and a classifier [33, 35].

In [36] PhishBlock was proposed as a hybrid tool that relied on lookup and a support vector machine classifier to check features derived from URL, text and linkage of visited websites. On the other hand, other researchers proposed information flow-based anti-phishing tools such as in [39] PhishGuard to submit bogus credentials during the user's login process and sent the actual credentials to identify phishing websites. Likely, Bogus Bitter proposed in [37], submitted a large number of bogus credentials along with actual user's credentials to nullify a phishing attack. And, PhishTester mitigated phishing websites that exploit cross site scripting vulnerabilities XSSVs of web browsers for malware distribution [38].

In addition, many industry researchers have released some anti-phishing tools such as Netcraft, and McAfee Site Advisor. Netcraft is produced by (netcraft.com, 2010), it assesses phishing site by trying to determine how old the registered domain of the visited website and it relies on a database maintained by a company [10, 15, 26, 27, 29]. Whilst, McAfee Site Advisor is a database-based anti-phishing tool that includes automated crawlers that browse websites and perform test for authenticity rating of the visited websites [20, 40, 41].

III. Issues and Future Trends

Based on the literature, almost researchers investigated issues behind the detection accuracy and computational cost of existing anti-phishing tools and conducted more researches to improve them towards obtaining optimum anti-phishing campaigns. However, they rarely addressed issues behind detection incapability of some notably effective anti-phishing tools against novel phishes which becomes a bottleneck of the existing anti-phishing campaign [3, 5, 11, 17, 18]. Phishers frequently change their behaviours and activities to avoid existing anti-phishing campaigns whenever they exploit novel phishes such as XSS-based, embedded object-based and new phishes hosted in any language-based websites that have not yet identified before by existing anti-phishing campaign. XSS-based phishes deploy XSSVs and obfuscated scripts for malware delivery. Embedded object based phishes imitate embedded components of web content such as Applets, Flash objects, ActiveX objects and advertising banners for advanced deceptions. On the other hand, newly emerged phishes can deploy some non-English language websites which have not yet been analyzed and identified for hostage [19, 22, 25, 42]. Based on the literature, Table II in Appendix reviews the detection capability of some notable anti-phishing tools in terms of language independence, XSS-based and embedded objects-based phishes.

In Table II of Appendix, the anti-phishing tool of [7] lacks leveraging XSS vulnerabilities of web browsers, and images, scripts, flash and ActiveX objects in the webpage source code for imitation and obfuscation. Most of the heuristics-based anti-phishing tools such as those proposed in [25, 31, 34], rarely leveraged novel phishes. Likely, hybrid based anti-phishing tools scarcely tolerate with novel phishes and phishes hosted in any language dependent website such as CANTINA$^+$ [35]. Furthermore, information flow-based anti-phishing tools such as those in [37-39], are detectible against any language-hosted phishing websites but they still can be bypassed by XSS-based and embedded objects-based phishes. Fig. 2 illustrates the state of the art comparison of anti-phishing tools.

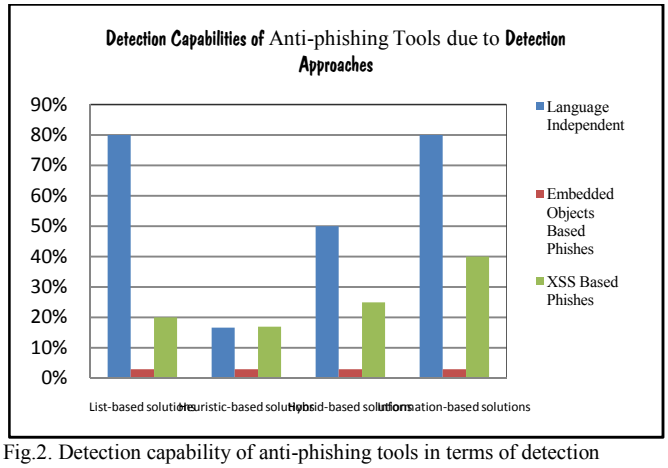


Fig.2. Detection capability of anti-phishing tools in terms of detection approaches and novel phishes.

Regarding to Appendix Table II and Fig. 2, almost surveyed anti-phishing tools fall shortly in detecting novel phishes. More importantly, they analyze phishing attacks by using design features and mechanisms that cannot leverage well those exploited by novel phishes and have not yet been identified. And they come up well with language dependent features like term frequency-inverse document frequency (tf-idf) features, text categorization and some language processing algorithms, and ready-made frequency lists of keywords that are best suited for some specific languages such as English language [18, 19, 21, 22, 25, 42]. Particularly, they utilize data sets consists of websites hosted English language rather than other languages. Thus, phishers can easily defeat them by exploiting these gaps and deliver new variants of attacks that have not yet been analyzed before. For example, heuristics-based anti-phishing tools mostly relied on term frequency-inverse document frequency (tf-idf) features and text categorization, which are language dependent features and mechanisms. Thus, they can effectively detect phishing websites made up with their own adapted heuristics only [18, 22, 42]. Similarly, hybrid-based anti-phishing tools lack of analyzing webpages made up of images, flash objects, applets, ActiveX objects and external hyperlinks. Furthermore, they rarely leverage obfuscated client side scripts that could be probably injected by phishers for malware delivery [22, 42].

With respect to these aforesaid issues, the detection capability of existing anti-phishing tools against novel phishes can be mainly considered along with detection accuracy as major concerns of research progress. We suggest that the ongoing research should focus in the facets of detection capability such as exploring new variants of features and deploying more sophisticated ones including embedded components, XSS-based features and client side scripting such as Java Scripts, PHP and ASP as well as unlimited keywords lists and language independent features which can be suitable for any natural language rather than English like eastern languages (Chinese and Arabic). In addition, new detective strategies with the aid of multifaceted computational science algorithms and techniques for content extraction and features similarity assessment should be emerged to leverage well novel phishes hosted in both websites and emails. At last, next wave of researches must be conducted to improve existing anti-phishing campaigns for wider scale detection of phishing attacks and provide essential factors to meet these issues.

IV. CONCLUSIONS

Concerns about novel phishes and the detection capability of the existing anti-phishing campaign have been arisen in recent years. And a continuing enrichment of the literature via wider objectives, theoretical and practical contributions is needed to meet cybersecurity requirements and financial indexes. More and new scenarios should be considered to deal with the novel activities of phishers and to reduce their risks. With the hope of stimulating researchers' interests and attention into the problem of detection capability against novel phishes, this research surveys the most up to date state of the art of anti-phishing campaigns and a large number of related work. In addition it attempts to address the recent gap of anti-phishing campaign that needs to bridge by describing and characterizing its elements.

Based on this survey, we reveal that the given issues fall into several major facets like features and mechanisms which could be developed for wider and effective detection of novel phishes. And there is still a long way to go on towards finding an optimum anti-phishing solution against all sophisticated phishes that can be probably exploited by phishers to bypass existing anti-phishing solutions.

ACKNOWLEDGMENT

The authors thank Universiti Teknologi Malaysia (UTM) for supporting this research.

REFERENCES

[1] M. Khonji, Y. Iraqi & A.Jones, "Phishing detection: a literature survey," Comm. Surveys & Tutorials, vol. 15, no. 4, pp. 2091–2121, 2013.

[2] M. He, S.J. Horng & P.Fan, "An efficient phishing webpage detector," Expert Systems With Applications, vol. 10, no. 38, pp. 12018–12027, 2011.

[3] A. Upadhyaya, "Design & development of a plug-in for a browser against phishing attacks," International Journal of Emerging Technology & Advanced Eng., vol. 2, no. 3, 2012.

[4] E.H. Chang, K.L. Chiew & S.N. Sze,"Phishing detection via identification of website identity," 2013 Int. Conf. IT Convergence Security (icitcs), pp. 1–4, 2013.

[5] Y. Lie, R. Xiao & J.Feng, "A semi-supervised learning approach for detection of phishing webpages," Optik-Int. J. for Light Electron Optics, vol. 14, no. 23, pp. 6027–6033, 2013.

[6] P. Likarish, E. Jung & D.Dunbar, "B-apt: Bayesian anti-phishing toolbar," ICC'08 Int. Conf. Comm., pp. 1745–1749, 2008.

[7] W. Han, "Using automated individual white-list to protect web digital identities," Expert Syst.With Applications, vol. 39, no. 15, pp. 11861–11869, 2012.

[8] L. Ma, "Detecting phishing emails using hybrid features," Symposia Workshops Autonomic Trusted Comput. Uic-atc'09, pp. 493–497, July 2009.

[9] Anti-Phishing Working Group, Phishing Archive, at http://www.antiphishing.org/phishing_archive.html

[10] G.S. Bindra, "Efficacy of Anti-phishing Measures and Strategies-A research Analysis, "World Academy Science, Eng. Technology, vol. 70, 2010.

[11] S. Pravin, "A phishing analysis of web based systems," 2011 Int. Conf. Comput. & Security, Proc. Communication,acm. , 2011.

[12] W. Kim, "The dark side of the Internet: Attacks, costs and responses," Inform. Syst., vol. 36, no. 3, pp. 675–705, 2011.

[13] H. Huang, S. Zhong & J.Tan, "Browser-side countermeasures for deceptive phishing attack," Ieee Fifth Int. Conf. Inform. Assurance Security Ias'09., vol. 1, pp. 352–355, Aug. 2009.

[14] H. Al-khateeb, "Security and usability in click-based authentication systems," Doctoral Dissertation, University of Bedfordshire, 2011.

[15] B. Wardman, "A series of methods for the systematic reduction of phishing," Doctoral Dissertation, University of Alabama, 2011.

[16] G. Gupta & J. Pieprzyk, "Socio-technological phishing prevention," Macquarie University, Research Online, 2011.

[17] A. San Martino & X. Perramon, "Phishing Secrets: History, Effects, Countermeasures," Ij Network Security, vol. 11, no. 3, pp. 163–171, 2010.

[18] S. Purkait, "Phishing counter measures and their effectiveness–literature review," Inform. Management & Comput. Security, vol. 5, no. 20, pp. 382–420, 2012.

[19] H. Shahriar, "Trustworthiness testing of phishing websites: a behavior model-based approach," Future Generation Comput. Syst., vol. 8, no. 28, pp. 1258–1271, 2012.

[20] R. Dhanalakshmi, "Detection of phishing websites and secure transactions," Int. J. Communication & Network Security (ijcns), vol. 1, pp. 15–21, 2011.

[21] S. Sheng, B. Wardman & C.Zhang, "An empirical analysis of phishing blacklists," Sixth Conf. Email Anti-spam (ceas), July 2009.

[22] R. Gowtham, I. Krishnamurthi & K.Kumar, "An efficacious method for detecting phishing webpage through Target Domain Identification," Decision Support Syst., 2014.

[23] M.G. Alkhozae & O.A. Maratfi, "Phishing websites detection based on phishing characteristics in the webpage source code," Int. J. Inform. Communication Technology Research., 2011.

[24] I. Jo, E.E. Jung & Y.H. Yeom, "Interactive Website Filter for Safe Web Browsing," J. Inform. Science & Eng., vol. 1, no. 29, 2013.

[25] W.D. Yu, "Phishcatch-a phishing detection tool," 33rd Annual Ieee Int. Comput. Software Applications Conf., 2009. Compsac'09, vol. 2, pp. 451–456, July 2009.

[26] M. Bhati, "Prevention Approach of Phishing on Different Websites," Int. J. Eng. Technology, vol. 2, no. 7, 2012.

[27] W. Chu, X. Guan & Z.Cai, "Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing URLs," 2013 Ieee Int. Conf. Comm. (icc), pp. 1990–1994, June 2013.

[28] S. Chaudhary, "Recognition of phishing attacks utilizing anomalies in phishing websites," Masters Dissertation, University of Tampere, 2012.

[29] J. Chhikara, "Phishing & Anti-Phishing Techniques: Case Study," Int. J. Advanced Research Comput. Science Software Eng., vol. 3, no. 5, pp. 458–465, 2013.

[30] A. Almomani, B. Gupta & E.Almomani, "A survey of phishing email filtering techniques," Ieee Comm. Surveys Tutorials, vol. 4, no. 15, pp. 1–21, 2013.

[31] C. Whittaker, "Large-Scale Automatic Classification of Phishing Pages," Nids, Mar. 2010.

[32] P. Prakash & R.R. Kompella, "PhishNet: predictive blacklisting to detect phishing attacks.," 2010 Ieee Proc. Infocom , pp. 1–5, Mar. 2010.

[33] Y. Zhang, "Cantina: a content-based approach to detecting phishing web sites," Proc. 16th Int. Conf. World Wide Web, Acm, pp. 639–648, May 2007.

[34] S. Gastellier-prevost, "Decisive heuristics to differentiate legitimate from phishing sites," IEEE 2011 Conf. Network Inform. Syst. Security (sar-ssi), pp. 1–9, May 2011.

[37] C. Yue & H. Wang, "BogusBiter: a transparent protection against phishing attacks," Acm Trans. Internet Technology, College William Mary, vol. 10, no. 2, 2010.

[38] H. Shahriar, "PhishTester: automatic testing of phishing attacks," Secure Software Integration and Reliability Improvement (ssiri), June 2010.

[39] Y. Joshi, "PhishGuard: A browser plug-in for protection from phishing," IEEE 2nd Int. Conf. Internet Multimedia Services Architecture Applications,imsaa 2008, pp. 1–6, Dec. 2008.

[35] G. Xiang, "CANTINA+: a feature-rich machine learning framework for detecting phishing web sites," ACM Trans. Inform. Syst. Security (tissec), vol. 2, no. 14, 2011.

[36] H.M. Fahmy, & S.A. Ghoneim, "PhishBlock: A hybrid anti-phishing tool," 2011 Ieee Int. Conf. Comm., Comput. Control Applications (ccca), pp. 1–5, Mar. 2011.

[40] N. Witte, "Rating the Authenticity of Websites," 16th Twente Student Conf. It , Jan 2012.

[41] R.B. Basnet, "Rule-based phishing attack detection," Int. Conf. Security Management (sam 2011), Las Vegas, Nv., 2011.

[42] R. Gowtham, "A comprehensive and efficacious architecture for detecting phishing webpages," Computers & Security, vol. 40, pp. 23–37, 2014.

APPENDIX

TABLE I. NOTABLE ANTI-PHISHING TOOLS WITH THEIR RELATIVE MERITS

Solution	**Year**	**Approach**	**Type**	**Contribution**	**Platform**	**Application level**
PhishGuard [39]	2008	Information flow	Plug-in	Website filter	Browser Independent	Client-side
B-APT [6]	2008	Whitelist	Toolbar	URL filter	Mozilla Firefox	Server-side
BogusBiter [37]	2010	Information flow	Toolbar	Website filter	Browser independent	Client-side
PhishTester [38]	2010	Information flow	Toolbar	Website filter	Internet Explorer7	Client-side
PhishCatch [25]	2010	Heuristics	Plug-in	Email filter	Browser Independent	Client-side
McAfee Site Advisor [13, 37]	2010	Hybrid	Extension	Website filter	McAfee Anti-virus	Client-Server
PhishNet [32]	2010	Blacklist	Toolbar	URL filter	Google	Client-side
PhishBlock [36]	2011	Hybrid	Toolbar	Website filter	Mozilla Firefox, Internet Explorer	Client-side
Google Toolbar [31]	2011	Heuristics	Toolbar	URL / Gmail filter	Google	Client-side
PhishShark [34]	2011	Heuristics	Toolbar	Website filter	Browser Independent	Client-side
CANTINA^{+} [35]	2011	Hybrid	extension	Website filter	Internet Explorer	Client-side
AIWL [7]	2012	Whitelist	Toolbar	URL filter	Browser Independent	Client-side

TABLE II. DETECTION CAPABILITY OF ANTI-PHISHING TOOLS IN TERMS OF XSS-BASED AND EMBEDDED OBJECTS-BASED PHISHES AS WELL AS LANGUAGE INDEPENDENCE

Related Work	**Brief Description**	**XSS-based phishes**	**Embedded objects-based phishes**	**Language independence**
AIWL [7]	It records legitimate websites URLs using Bayesian filter	Yes	No	Yes
PhisNet[32]	It maintains blacklist of phishing URLs using TLD and DNs features	No	No	Yes
CANTINA+ [35]	Extract features of webpage identity and compare them with the current domain using search engine.	No	No	No
PhishShark [34]	Identifies phishness and legitimacy of websites using twenty heuristics	No	No	No
PhishGuard [39]	It Submits fake credentials before and after actual user's credentials.	Yes	No	Yes
BogusBiter[37]	It sends bogus credentials when a webpage is detected as phishing to avoid information leakage.	No	No	Yes
PhishTester [38]	Identifies phishing websites by using FSM and several features.	Yes	No	Yes
PhishBlock [36]	It is based on both lookup and a SVM classifier that checks features derived from websites URL, text and linkage.	No	No	No
Google Toolbar [31]	It classifies phishing emails and webpages using classifier and Google's blacklist	No	No	No
PhishCatch [25]	It analyzes phishing emails using heuristics	No	No	No
McAfee Site Advisor [38]	It maintains list of website's safety ratings.	Yes	No	Yes
B-APT [6]	It identifies phishing websites by using Bayesian filter and DOM tree.	No	No	Yes