Int. J. Advance Soft Compu. Appl, Vol. 8, No. 1, March 2016
ISSN 2074-8523

New Hybrid Features for Phish Website Prediction

Hiba Zuhair[1], Ali Selamat[2,3], and Mazleena Salleh[3]

[1]Al-Nahrain University, Baghdad, Iraq
e-mail: hiba.zuhair.pcs2013@gmail.com,
[2]Department of Computer Science,
Faculty of Computing, Universiti Teknologi Malaysia,
81310 UTM Johor Bahru, Johor, Malaysia;
e-mail: mazleena@utm.my
[3]UTM-IRDA Center of Excellence,
Universiti Teknologi Malaysia,
81310 UTM Johor Bahru, Johor, Malaysia
e-mail: aselamat@utm.my

Abstract

Phishing is a serious threat to the web economy and the Internet communication, because phishers put both users and organizations at risk of identity theft and financial losses. Phishers continually exploit new sophisticated features to impersonate legitimate web pages, modify their components and host their phishes. Furthermore, the prediction susceptibilities of features that were previously investigated become a key challenge for discriminating the evolving phishes. Accordingly, this paper investigated the prediction susceptibility of 58 hybrid features. It was observed that the investigated features were highly exploited in the content and hosted the URLs of phish webpages. The prediction susceptibility of the proposed features was experimentally examined in the suspected webpages using the SVM machine learning classification technique. The results revealed that the introduced features could be considered as potentially predictive ones and they could be utilized in the upcoming research to improve phishing detection approaches.

Keywords: *Hybrid features, Phishing, Prediction Susceptibility.*

1 Introduction

In the last decade, cyberspace has shown a rapid expansion of phishing. Phishers try to target users and enterprises to access their sensitive information. They imitate legitimate websites with some deceptive features to build their phishes [1-3]. Moreover, they continually evolve phishes by exploiting more sophisticated features in different feature spaces, such as webpage URLs and content. Thus, they can circumvent the existing phish detect approaches, causing more potential risks and monetary losses [4, 5]. Most of the literature focus on methods of surviving phish attacks, hosted in webpages and the ways to improve the existing phishing detective approaches, such as the list-based, heuristics, hybrid and information flow-based methods [1, 2, 6, 7]. The hybrid detective approaches somewhat outperform other approaches due to the use of classifiers and multiple types of features, i.e. hybrid features. However, in the course of this approach, some related issues have emerged, such as the evolutionary phishing features, so that the recently deployed ones are not effective in handling them. This, in turn, has degraded the performance of the detective phishing approach against the rapid growth and distribution of phish webpages over the Web [3, 6, and 7]. As such, exploring more predictive features continually will help improve effective phishing detection. For this purpose, this study sought for newly exploited features by the phishers and experimentally investigated them in terms of the phishing prediction susceptibility.

Thus, by examining 58 hybrid features from the webpage URL and content, the webpage URLs were extracted and investigated through certain computational strategies, such as the features selection and classification. The scope of this work is limited to the experimental investigation of numerous new features for phishing detection. The computational strategies related to the features selection and machine learning classification were kept constant, but they can be improved for the recommended features in the future research. The rest of this paper is organized as follows. Section 2 presents a background of the previously proposed features and detective strategies in the context of existing related studies. Section 3 presents the newly introduced hybrid features. In Section 4, the assessment strategy and experiments are explained. Section 5 discusses the results in the context of the experiments. Finally, conclusions and future implications are presented in Section 6 to give insight on the obtained results.

2 Related Works

Various phishing detective approaches have been proposed by researchers in the recent years to mitigate the increased phishing susceptibility. In general, these approaches can be classified into white lists of famous trustworthy URLs, black lists of valid phish URLs, heuristics and rule-based approaches, information flow and hybrid approaches. However, most of them have several shortcomings in leveraging features that are continually exploited by the phishers. Table 1 summarizes the previously deployed features for phishing detection and characterizes them into the webpage content features, the URL features and the online features or third party features based on their nature and the webpage parts [8, 9, and 28]. These features are varied in their prediction susceptibility to phishing so that each feature may have negative or positive effects on the performance of phishing detective approach.

Evolving phish webpages contain various hybrid features; some of which were rarely considered and defined by the previous works. Today, phishers exploit deceptive features in target webpages to hide some links for users' redirection to their own fake webpages. Moreover, they obfuscate the client-side scripting components, such as JavaScript, PHP and ASP. Moreover, they modify some applets, Flash objects and ActiveX controls in the source file of their targets to submit their cookies and fake advertisements through the web banners. They also target the URLs of webpages presented in any language rather than English, e.g. Chinese e-business webpages. Such deceptive features enable phishers to install suspicious, malicious and spy codes into the client's computer for further damages and create multiple replicas of their targets for pharming purposes, i.e. redirecting as many visitors as possible to the same fake website. Also, they exploit host URLs of non-English webpages to bypass those phishing detective approaches that have not yet identified them. Today, all these issues are big challenges in mitigating phishing over the web, because the existing phishing detection approaches cannot predict such phishing features [3, 5, 17, 22, 36, 37, 40, 43, and 46].

3 Investigation of Features and Their Prediction Susceptibilities

Based on the aforementioned literature, an experimental strategy was adopted to explore new important deceptive features as shown in Fig. 1. Total of 58 hybrid features were explored manually in this work. In addition, their prediction susceptibility against the above-mentioned phishing deceptions was experimentally investigated. The examined features belong to two different features spaces: the webpage content and the URL. The first feature space contained 48 features that are mostly cross-site scripting and embedded objects.

Table 1: Summarization of the features explored by the literatures

Feature Space	**Feature Type**	**Features**
URL Features [26-30, 47]	Structural	IP address, length of hostname length, dots in hostname, dashes in hostname, long hostname, shortened URL, certain characters (“@”, “//”), misspellings and derivation of domain names (paypal1 or paypla), port number, domain’s life, encoded URL (ASCII, Hexa, Oct), Abnormal anchors, RURLs (abnormal request URLs).
	Lexical	Tokens (confirm, banking, secure, ebayispi, webscr, login, signin), HTTP instead of HTTPS
	Brand-name	Names of brands of targeting sites and companies (eBay, Paypal, sulake, facebook, orkut, Santander,mastercard, warcraft, visa, bradesco).
Webpage content [31, 34, 36-38, 40-43]	Linkage Structure and Source code	Coupled In/back links, out links within body text Components of DOM tree, title of the webpage, description fields of Meta, tags of HTML document, files of Cascading Style Sheet (CSS), SQL injections, scripts, input, text, password, and hidden fields Internal and external hyperlinks in <a herf=” “><a/>, buttons and their actions, Illegal pop-up windows, unfamiliar English, email functions.
	Identity	WHOIS data (registrant, registration, expiration dates)
	Visual clues	Images, logos, overall visual layout, styles, blocks, key regions, rendering, flash objects, background color, font family, text alignment, and line spacing.
Online features [3, 17, 18, 22, 23, 46]	Domain Address	Age and history of Domain. History of URL address, ranking the webpage in search engine results, certificate state.
	Others	Number of visitors, number of links to the website

Whereas the second feature space contained 10 URL features as presented in Tables 2 and 3, respectively.

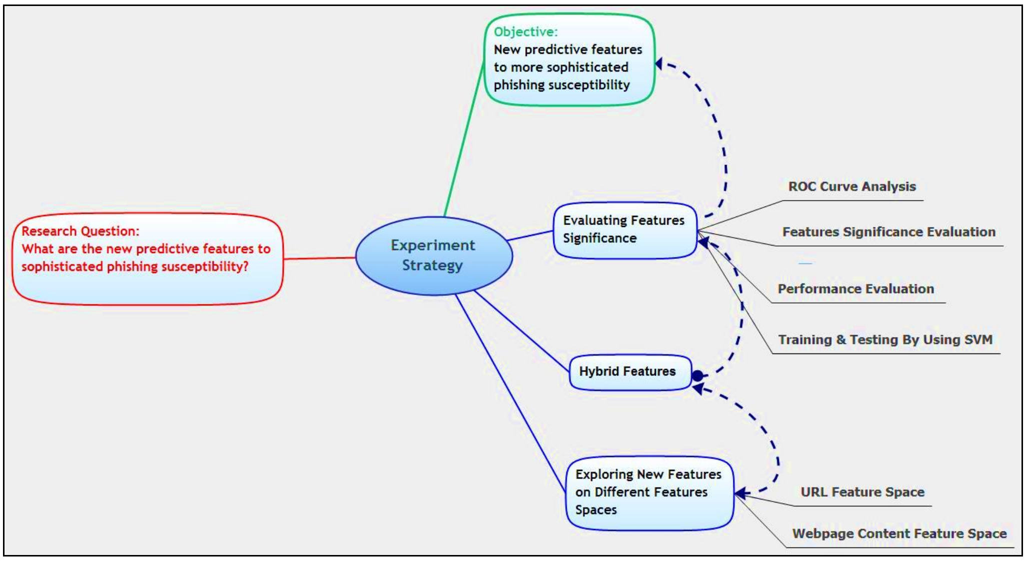
Objective:
New predictive features to more sophisticated phishing susceptibility
Research Question:
What are the new predictive features to sophisticated phishing susceptibility?
Experiment Strategy
Evaluating Features Significance
ROC Curve Analysis
Features Significance Evaluation
Performance Evaluation
Training & Testing By Using SVM
Hybrid Features
Exploring New Features on Different Features Spaces
URL Feature Space
Webpage Content Feature Space

Fig. 1: Experiment strategy in terms of research question, research objective and experimental steps

It is highly important to investigate whether or not the explored features can be effective in predicting phishing susceptibility. The examined webpage is signified by the standard document representation that is usually used for text classification. Each examined document j is represented as a feature vector $D_j = \{d_{j,1}, \quad d_{j,2}, \ldots, \quad d_{j,n}\}$, where n is the number of features, $d_{j,i}$ indicates the feature itself as a numeric value so that $0 \leq d_{j,i} \leq 1$ [9, 14, 25]. Then, to state the significance of new hybrid features, the given feature matrix F was trained through the Support Vector Machine (SVM) classifier. The SVM classifier is the most commonly used classifier to obtain the optimal separating hyper plane between two classes [6-8, 10, 11, 24, 32, and 33]. It guarantees the lowest level of error rate because of its generalization ability and capacity of handling high dimensional feature space. Furthermore, the SVM classifier produces two output classes [8,12, 32], which are represented by the label of +1 and -1 as follows:

F denotes all the webpages in the dataset, i.e. a multi-dimensional features matrix consisting of a set of feature vectors, so that $F = \{F_1 \quad F_j \quad F_{|F|}\}$ and F_j is the feature vector of each webpage as $F_j = \{f_{j,1} \quad f_{j,i} \quad f_{j,|F_j|}\}$, where $|F|$ and $|F_j|$ are the number of feature vectors and features in each feature vector, respectively.

Table 2: Extracted webpage content features

Embedded Objects and Links Features		Cross Site Scripting Features	
Index	Features	Index	Features
F1	Number of Scripting.FileSystemObjec	F25	JavaScript scripts length
F2	Number of Excel.Applicati	F26	Number of functions’ calls in java scri
F3	Presence of WScript.shell	F27	Number of script lines in java scripts
F4	Presence of Adodb.Stream	F28	Script line length in java scripts
F5	Presence of Microsoft.XMLDOM	F29	Existence of long variable names in ja scripts
F6	Number of <embed>	F30	Existence of long function names in ja scripts
F7	Number of <applet>	F31	Number of fromCharCode()
F8	Number of Word.Applicati	F32	Number attachEvent()
F9	link length in <embed>	F33	Number of eval()
F10	Number of <iframe>	F34	Number of escap()
F11	Number of <frame>	F35	Number of dispacthEvent()
F12	Out-of-place tags	F36	Number of SetTimeout()
F13	Number of <form>	F37	Number of exec()
F14	Number <input>	F38	Number of pop()
F15	Number of MSXML2.XMLHTTP	F39	Number of replaceNode()
F16	Frequent <head>, <title>, <body>	F40	Number of onerror()
F17	<meta index.php?Sp1=>	F41	Number of onload()
F18	“Codebase” attribute in <object>	F42	Number of onunload()
F19	“Codebase” attribute in <applet>	F43	Number of <script>
F20	“href” attribute of <link>	F44	frequent<div onClick=window.open()’
F21	Number of void links in <form>	F45	Number of <script>
F22	Number of out links	F46	Number of MSXML2.XMLHTTP
F23	Number of <form> in java scripts	F47	Number of onerror()in javascripts
F24	Number <input> in java scripts	F48	Number of SetInterval()

Table 3: Extracted URL features

Index	Webpage’ URL Features	Index	Webpage’ URL Features
F49	Multiple TLD	F54	Typos in Base name
F50	Brandname in hostname	F55	Long domain name
F51	Special symbols in URL	F56	Misleading subdomain
F52	Coded URL	F57	Number of dots in URL
F53	IP address instead of domain nan	F58	Path domain length

Then, $f_{j,i}$ is the value of each i^{th} feature of j^{th} feature vector F_j, where$0 \leq f_{j,i} \leq 1$, $i = 1, 2, 3, \ldots, |F_j|$ and$j = 1,2,3,\ldots,|F|$, given that $F = \{F_j\}_{j=1}^{|F|}$ is a set of $|F|$ training feature vectors or alternatively, the M-dimensional feature matrix. Each F_j is labelled by $y_j \in \{1, -1\}$ with $y_j = 1$ and$y_j = -1$, which indicates the membership of F_j in the class 1 and class 2 through Equation 1 [6, 14].

$$f(x) = \sum_j \alpha_j \gamma_j K(F', F_j) + b \quad (1)$$

Where $\boldsymbol{\alpha_i}$ and b are obtained by a quadratic algorithm, F' is the unlabelled webpage and $\boldsymbol{F_i}$ is the feature vector of a training webpage. The function $\boldsymbol{K(F', F_i)}$ maps the space of input webpage to higher dimensions, where training webpages in the dataset are learned individually.

4 Results and Discussion

4.1 Experimental Setup

A preliminary set of real world webpages, 500 living phishing webpages and 500 valid legitimate webpages were downloaded in 60 days from September to November 2014. Specifically, the phishing pages were downloaded from two publically available sources; the Phish Tank and the Castle Cops archives. The Alexa’s top sites archive was used as the source of legitimate webpages. The collected webpages were different from those of the most targeted financial organizations, homepage, and login functionalities. Experimentally, the SVM classifier was trained four times on three matrices belonging to three feature spaces. The first feature space represented the webpage content feature space. Unlikely, the second features space contained features extracted from the webpage URL. The third feature space contained a combination of the former feature spaces to represent the hybrid feature space. Each feature matrix is a multi-dimensional matrix, where each row represents a feature vector extracted from an examined webpage. A feature vector contains the values of all the features extracted from an examined webpage and its label appears in the first column. Furthermore, the values of those features are either binary or numeric values. The binary features were computed as the union of their corresponding features, while the numeric features were combined by taking the smallest value of the corresponding features. First, all of the feature matrices were trained and tested using the SVM over the collected dataset. To implement the SVM, a machine learning tool from the Waikato Environment for Knowledge Analysis (WEKA) was used. Then, performance of the SVM over the collected dataset was evaluated using the newly introduced hybrid features according to the experimental results. The performance represented the percentage of correctly examined webpages against the phishing susceptibility over the total number of webpages included in the dataset.

4.2 Results

To demonstrate the significance of hybrid features proposed in this study, the SVM classification was used to explore feature spaces and some formerly used measures. These measurements included the TP, FP, FN, Precision, Recall, and the F1-measure as well as the AUC values under the ROC curve [23, 32, 33, 35, and 46]. The TP or True Positive indicates the rate of correctly classified phish instances. The FP (False Positive) refers to the rate of wrongly classified legitimate instances as the phishing ones. The FN (False Negative) indicates the wrongly labelled phish instances as legitimate ones [23, 32, 33, and 46]. Results of the TP, the FP and the FN are illustrated in Fig. 2 and Fig. 3, respectively. Each of the Precision, Recall and F-measure was computed using the parameters of TP, FP and FN [23, 24, 32, and 33]. The maximal value of precision states the maximal positive webpages that are classified. However, the maximal recall value denotes minimal prediction error. Then, the F-measure was used to harmonically compute the mean of both aforementioned measurements, which denotes the initial phishes indication of the extracted features. Equations 2, 3 and 4 describe these measurements [6-8, 10, 11, 23, 24, 32, and 33]. Fig. 4 illustrates the features evaluation with respect to the above-mentioned measurements using the SVM classifier.

$$Precision = \frac{|TP|}{|TP|+|FP|} \tag{2}$$

$$Recall = \frac{|TP|}{|TP|+|FN|} \tag{3}$$

$$F - measure = 2\times\frac{Precisio\times Recall}{Precision+Recall} \tag{4}$$

On the basis of the classifier's prediction for each examined webpage in the dataset, the area under the ROC curve (AUC) was set for each feature. Next, each feature was ranked in the form of its weight value and the features were grouped according to the weight values, positive features in weight and negative features in weight, respectively. Then, the features in each group were sorted in a descending order. Calculation of the AUC value is presented in Equation 5 [35, 46] as follows:

$$AUC = \frac{1}{MN}\sum_{j=1}^{M}(S_j - J) \qquad (5)$$

Where S_j is the rank of each j^{th} feature in each group and $S_j - J$ is the number of positive features before the negative features in weight [35, 46]. The results of the computing area under the ROC curve in terms of the features space and AUC value is illustrated in Fig. 5.

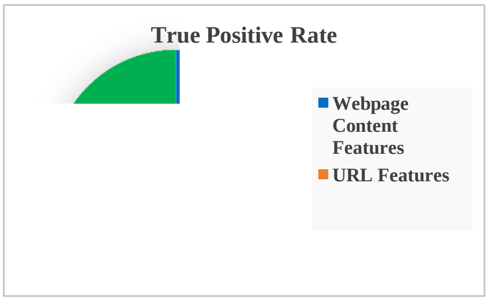

Fig.2: Percentages of TP in terms of the features category.

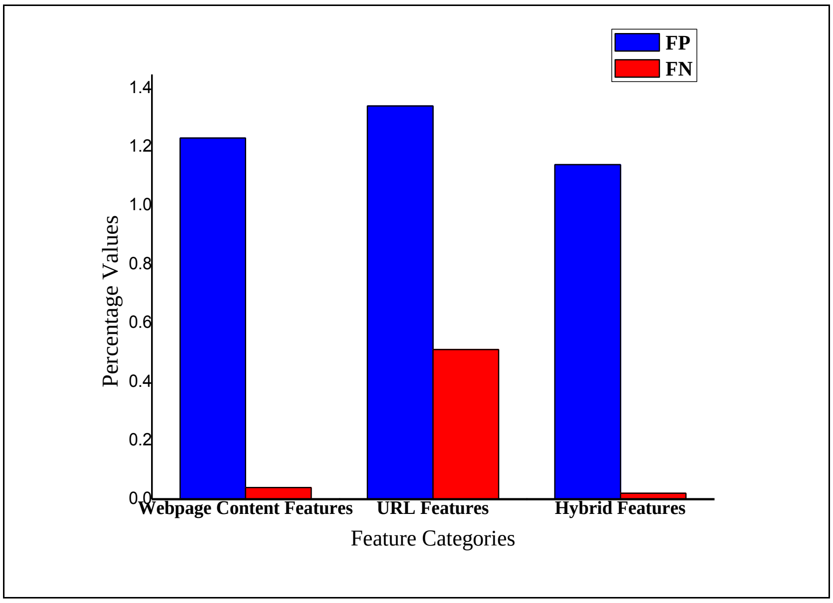

Fig.3: Prediction sensitivity in terms of the category of features space, and percentages of FP and FN.

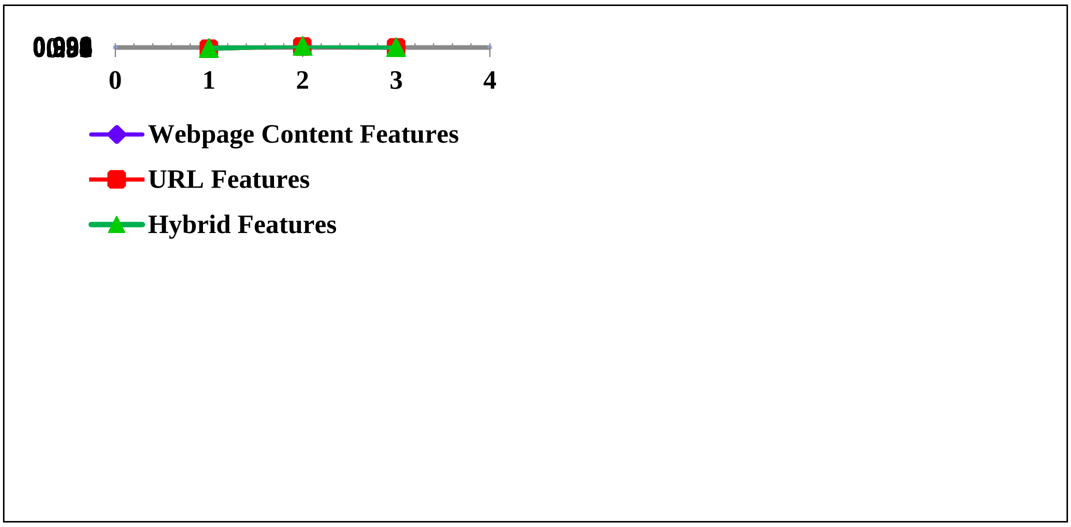

Fig. 4: Prediction merits in terms of features space, Precision, Recall and F-measure.

4.3 Discussion

The aforementioned experimental results and evaluations revealed that based on the used features space and collected dataset, the results of the TP rate had the higher accuracy under the SVM machine learning algorithm as plotted in Figure 2. However, the results plotted in Figure 3 show that they achieved lower sensitivity to the FP and substantially lower FN as compared to other tested features. This is because the investigated features could handle a large dataset with various features belonging to multiple features spaces. On the other hand, the test results plotted in Figure 4 show that the proposed hybrid features performed the best under the SVM classifier based on all the three performance measurements (Precision, Recall and F-measure). Therefore, these can effectively maximize the prediction susceptibility of phishing detective approach against phishing as compared to other previous features. In addition, the statistical results illustrated in Figure 5 reveal that 58 features vary in their susceptibility to predict phishing. For instance, the features F1, F2, F3, F18, F49 and F52 outperform all other with the AUC values of 0.8844, 0.8829, 0.8772, 1.0000, 0.9624 and 0.8226, respectively. However, the features including F9, F11, F15, F51, F53 and F56 perform lower with the AUC values of 0.6391, 0.6428, 0.6362, 0.4537, 0.5621 and 0.2548. This implies that phishing prediction of these features under the SVM classifier is random.

To sum up, the findings show that new phishing features could possibly provide valuable trends of investigation for detecting more sophisticated phishes. This will help in circumventing the phishers’ attempts to bypass the existing phishing detective approaches.

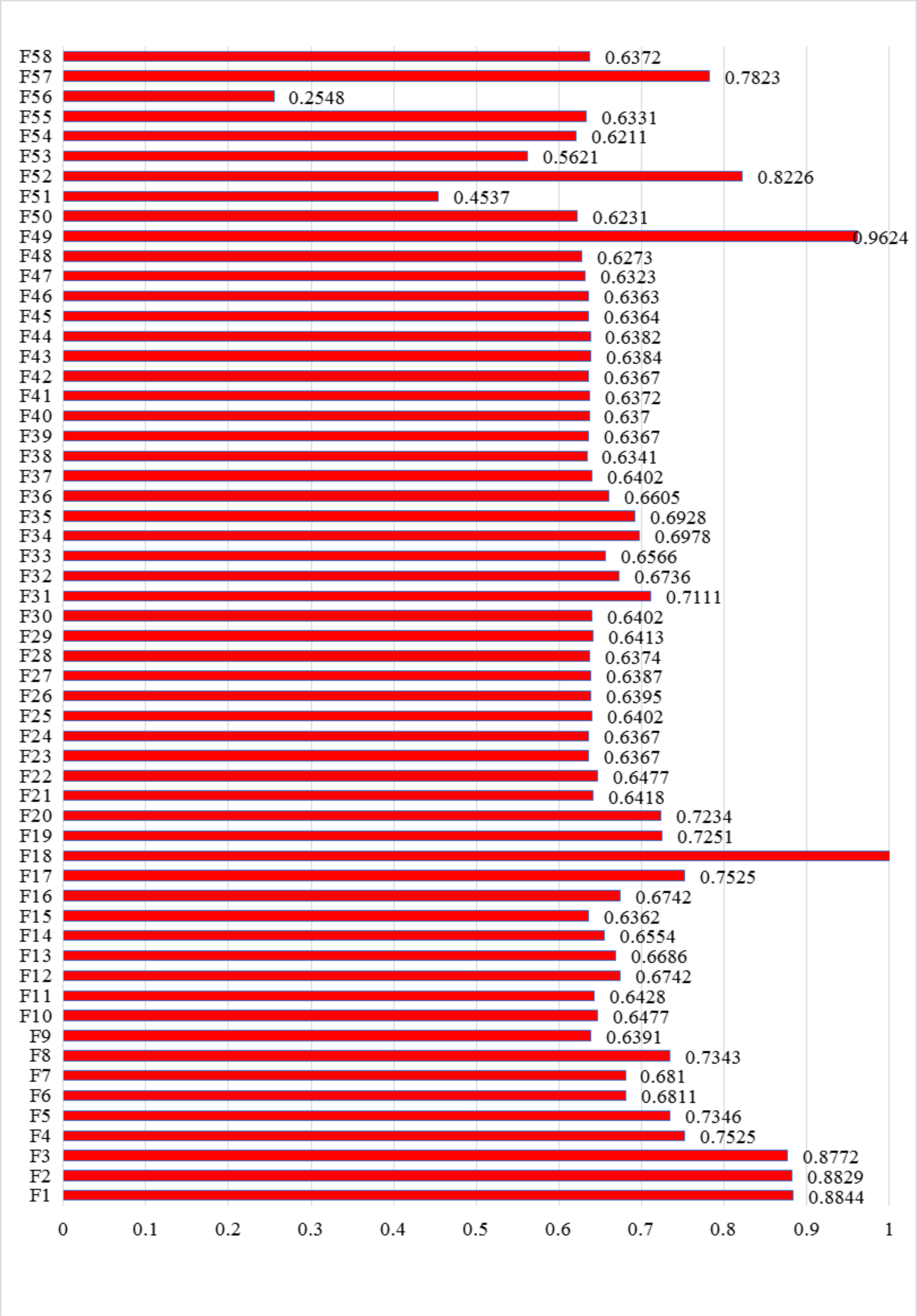

Fig. 5: Area under ROC curve in terms of individual features and their AUC values.

Consequently, almost all of the introduced features are highly expected to be exploited by the phishers owing to their functionalities in modification, imitation, redirection , injection of codes and links to obtain user's confidential information. These features can be assigned to more than one type of phish. Furthermore, accuracy and sensitivity tests indicate effects of the reduced overlap and increased assignment of hybrid features to more than one phish pattern. However, major issues that should be considered are the computation time and the cost. Even though the used SVM classifier is highly effective, using 58 hybrid features may negatively affect the execution time of extraction, training and testing over large datasets. Further improvement is needed to reduce the dimensionality of the introduced feature space as well as the complexity and the time of prediction over large datasets and for real life applications. Thus, single or multiple use of the aforementioned features are recommended against the sophisticated phishing attacks. Also, the best combination of features should be taken into account by using feature selection techniques to help in improving the overall performance of detection.

5 Conclusions

This work introduced 58 hybrid features for the effective prediction susceptibility on advanced phishing attacks. A supervised machine learning technique and some commonly used performance measurements were used to assess the introduced features through experiments. Two features spaces were explored to extract hybrid features. Then, these features were trained and tested using the SVM classifier to evaluate their prediction susceptibility with respect to their classification performance against the phish webpages over a collected dataset. According to the findings, the proposed hybrid features can contribute to high prediction susceptibility and yield accurate detection results compared to the previously utilized features in the literature. The scope of the present work is limited to introducing new hybrid features and precise assessment of their prediction susceptibilities against some kinds of emerging phishes. Thus, the future work should continually target major improvements aimed at reducing the time and complexity of features processing. Additionally, the detection approach should be improved through techniques such as the machine learning classification for the best prediction susceptibility in real life situations

References

[1] Khonji, M., Iraqi, Y. and Jones, A. 2013. Phishing detection: a literature survey, Communications Surveys & Tutorials, IEEE, vol. 15, (2013), pp. 2091-2121.

[2] Purkait, S. 2012. Phishing counter measures and their effectiveness–literature review, Information Management & Computer Security, vol. 20, (2012), pp. 382-420.

[3] Ramesh, G., Krishnamurthi, I. and Kumar, K. 2014. An efficacious method for detecting phishing webpages through target domain identification, Decision Support Systems, vol. 61 (2014), pp. 12-22.

[4] Tonge, A. M. and Chaudhari, S. R. Phishing Susceptibility and Anti-Phishing Security Strategies-Literature Review.IJSR.

[5] Sheeram, V., Suban, M., Shanthi, P. and Manjula, K. 2010. Anti-phishing detection of phishing attacks using genetic algorithm, In Proceedings of IEEE International Conference on Communication Control and Computing Technologies ICCCCT (2010), pp. 447-450.

[6] He, M., Horng, S.-J., Fan, P., Khan, M. K., Run, R.-S., Lai, J.-L. 2011. An efficient phishing webpage detector, Expert Systems with Applications, vol. 38 (2011), pp. 12018-12027.

[7] Li, Y., Xiao, R., Feng, J. and Zhao, L. 2013. A semi-supervised learning approach for detection of phishing webpages, Optik-International Journal for Light and Electron Optics, vol. 124 (2013), pp. 6027-6033.

[8] Barraclough, P., Hossain, M., Tahir, M., Sexton, G. and Aslam, N. 2013. Intelligent phishing detection and protection scheme for online transactions, Expert Systems with Applications, vol. 40 (2013), pp. 4697-4706.

[9] Mohammad, R. M., Thabtah, F. and McCluskey, L. 2012. An assessment of features related to phishing websites using an automated technique, In Proceedings of International Conference on Internet Technology And Secured Transactions (2012), pp. 492-497.

[10] Islam, R. and Abawajy, J. 2013. A multi-tier phishing detection and filtering approach, Journal of Network and Computer Applications, vol. 36 (2013), pp. 324-335.

[11] Sadi, M. S., Khan, M. M. R., Islam, M. M., Srijon, S. B. and Mia, M. M. H. 2012. Towards detecting phishing web contents for secure internet surfing, In Proceedings of International Conference on Informatics, Electronics & Vision ICIEV (2012), pp. 237-241.

[12] Likarish, P., Jung, E., Dunbar, D., Hansen, T. E. and Hourcade, J. P. 2008. B-apt: Bayesian anti-phishing toolbar, In Proceedings of IEEE International Conference on Communications ICC'08 (2008), pp. 1745-1749.

[13] Han, W., Cao, Y., Bertino, E. and Yong, J. 2012. Using automated individual white-list to protect web digital identities, Expert Systems with Applications, vol. 39 (2012), pp. 11861-11869.

[14] Whittaker, C., Ryner, B. and Nazif, M. 2010. Large-Scale Automatic Classification of Phishing Pages, In Proceedings of NDSS (2010).

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

[16]Yu, W. D., Nargundkar, S. and Tiruthani, N. Phishcatch-a phishing detection tool," In Proceedings of 33rd Annual IEEE International Computer Software and Applications Conference COMPSAC'09, pp. 451-456.

[17]Gastellier-Prevost, S. , Granadillo, G. G. and Laurent, M. 2011. Decisive heuristics to differentiate legitimate from phishing sites, In Proceedings of Conference on Network and Information Systems Security (SAR-SSI), pp. 1-9.

[18]Xiang, G., Hong, J., Rose, C. P. and Cranor, L. 2011. CANTINA+: a feature-rich machine learning framework for detecting phishing web sites, ACM Transactions on Information and System Security (TISSEC), vol. 14 (2011), p. 21.

[19]Fahmy, H. M. and Ghoneim, S. A. 2011. PhishBlock: A hybrid anti-phishing tool, In Proceedings of International Conference on Communications, Computing and Control Applications (CCCA), 2011, pp. 1-5.

[20]Joshi, Y. , Saklikar, S., Das, D. and Saha, S. 2008. PhishGuard: a browser plug-in for protection from phishing, In Proceedings of 2nd International Conference on Internet Multimedia Services Architecture and Applications (IMSAA 2008), pp. 1-6.

[21]Yue, C. and Wang, H. 2010. BogusBiter: A transparent protection against phishing attacks, ACM Transactions on Internet Technology (TOIT), vol. 10 (2010), p. 6.

[22]Shahriar, H. and Zulkernine, M. 2010. PhishTester: automatic testing of phishing attacks, In Proceedings of Fourth International Conference on Secure Software Integration and Reliability Improvement (SSIRI), pp. 198-207.

[23]Shahriar, H. and Zulkernine, M. 2012. Trustworthiness testing of phishing websites: A behavior model-based approach, Future Generation Computer Systems, vol. 28 (2012), pp. 1258-1271.

[24]Lakshmi, V. S. and Vijaya, M. 2012. Efficient prediction of phishing websites using supervised learning algorithms, Procedia Engineering, vol. 30 (2012), pp. 798-805.

[25]Wang, H., Zhu, B. and Wang, C. 2012. A Method of Detecting Phishing Web Pages Based on Feature Vectors Matching, Journal of Information and Computational Systems, vol. 9 (2012), pp. 4229-4235.

[26]Nguyen, L. A. T., To, B. L., Nguyen, H. K. and Nguyen, M. H. 2013. Detecting phishing web sites: A heuristic URL-based approach, In Proceedings of International Conference on Advanced Technologies for Communications (ATC2013), pp. 597-602.

[27]Zhang, J. and Wang, Y. 2012. A real-time automatic detection of phishing URLs, In Proceedings of 2nd International Conference on Computer Science and Network Technology (ICCSNT), pp. 1212-1216.

[28]Basnet, R. B., Sung, A. H. and Liu, Q. 2011. Rule-based phishing attack detection, In Proceedings of International Conference on Security and Management (SAM 2011), Las Vegas, NV.

[29]Basnet, R. B. and Sung, A. H. 2012. Mining Web to Detect Phishing URLs, In Proceedings of 11th International Conference on Machine Learning and Applications (ICMLA), pp. 568-573.

[30]Kordestani, H. and Shajari, M. 2013. An entice resistant automatic phishing detection, In Proceedings of 5th Conference on Information and Knowledge Technology (IKT), pp. 134-139.

[31]Alkhozae, M. G. 2011. Phishing websites detection based on phishing characteristics in the webpage source code, International Journal of Information and Communication Technology Research.

[32]Huang, H. Qian, L. and Wang, Y. 2012. A SVM-based technique to detect phishing URLs, Information Technology Journal, vol. 11 (2012), pp. 921-925.

[33]Zhuang, W., Jiang, Q. and Xiong, T. 2012. An intelligent anti-phishing strategy model for phishing website detection, In Proceedings of 32nd International Conference on Distributed Computing Systems Workshops (ICDCSW), pp. 51-56.

[34]Uzun, E., Agun, H. V. and Yerlikaya, T. 2013. A hybrid approach for extracting informative content from web pages, Information Processing & Management, vol. 49 (2013), pp. 928-944.

[35]Olivo, C.K., Santin, A.O. and Oliveira, L.S. 2011. Obtaining the threat model for e-mail phishing, Applied Soft Computing.

[36]Rajalingam, M., Alomari, S. A. and Sumari, P. 2012. Prevention of Phishing Attacks Based on Discriminative Key Point Features of WebPages, International Journal of Computer Science and Security (IJCSS), vol. 6 (2012), p. 1.

[37]Bhati, M. and Khan, R. 2012. Prevention Approach of Phishing on Different Websites, International Journal of Engineering and Technology, vol. 2 (2012).

[38]Ramya, K. R. , Priyanka, K. , Anusha, K., Devi, C. J. and Prasad, Y. S. An Effective Strategy for Identifying Phishing Websites using Class-Based Approach.

[39]Liu, Y. and Zhang, M. 2012. Financial websites oriented heuristic anti-phishing research, In Proceedings of IEEE 2nd International Conference on Cloud Computing and Intelligent Systems (CCIS), pp. 614-618.

[40]Sanka, K. and Suresh, B. A New Framework for Thwarting Phishing attacks based on Visual Cryptography.

[41]Mayuri, A. and Tech, M. 2012. Phishing Detection based on Visual-Similarity, In Proceedings of International Conference on Networks and Cyber Security, p. 5.

[42]Witte, N. 2012. Rating the Authenticity of Websites, (2012).

[43]Chang, E. H., Chiew, K. L. Sze, S. N. and Tiong, W. K. 2013. Phishing Detection via Identification of Website Identity, In Proceedings of International Conference on IT Convergence and Security (ICITCS), pp. 1-4.

[44]Fu, L., Meng, Y., Xia, Y. and Yu, H. 2010. Web content extraction based on webpage layout analysis, In Proceedings of Second International Conference on Information Technology and Computer Science (ITCS), pp. 40-43.

[45]Pan, Y. and Ding, X. 2006. Anomaly based web phishing page detection, In Proceedings of 22nd Annual Computer Security Applications Conference (ACSAC'06), pp. 381-392.

[46]Gowtham, R. and Krishnamurthi, I. 2014. A comprehensive and efficacious architecture for detecting phishing webpages, Computers & Security, vol. 40 (2014), pp. 23-37.

[47]Khonji, M., Iraqi, Y. and Jones, A. 2011. Lexical URL analysis for discriminating phishing and legitimate websites, In Proceedings of the 8th Annual Conference Collaboration on Electronic messaging, Anti-Abuse and Spam, ACM.