A Novel Neuro-Fuzzy Approach for Phishing Identification

Luong Anh Tuan Nguyen and Ba Lam To and Huu Khuong Nguyen and Chuan Pham and Choong Seon Hong

Abstract— **Together with the growth of Internet, e-commerce transactions play an important role in the modern society. As a result, phishing is a deliberate act by an individual or a group of people to steal personal information such as password, banking account, credit card information, etc. Most of these phishing web pages look similar to the real web pages in terms of website interface and uniform resource locator (URL) address. Many techniques have been proposed to identify phishing websites, such as Blacklist-based technique, Heuristic-based technique, etc. However, the number of victims has been increasing due to inefficient protection technique. Neural networks and fuzzy systems can be combined to join its advantages and to cure its individual illness. This paper proposed a new neuro-fuzzy model without using rule sets for phishing identification. Specifically, the proposed technique calculates the value of heuristics from membership functions. Then, the weights are trained by neural network. The proposed technique is evaluated with the datasets of 11,660 phishing sites and 10,000 legitimate sites. The results show that the proposed technique can identify over 99% phishing sites.**

I. INTRODUCTION

The word ”phishing” is produced from the word ”fishing”. Phishers, creating phishing sites, use a number of techniques to fool their victims, including email messages, instant messages, forum posts, phone calls and social networking. With these activities of phishing, it causes severe economy loss all over the world. According to a study by Gartner [1], 57 million US Internet users have identified the receipt of email linked to phishing scams and about 2 million of them are estimated to have been tricked into giving away sensitive information. Meanwhile, phishing sites are also growing rapidly in quality and quantity. Therefore, the risk of stealing user information is extremely high. Because of these reasons, identifying phishing problem is very urgent, complex and extremely important problem in modern society. Recently, there have been many studies that against phishing based on the characteristics of site, such as URL of website, content of website, combining both the website URL and content, source code of website or interface of website, etc. However, each of studies has its own strengths and weaknesses. There is still not a sufficient method. In this paper, a new approach is proposed to identify the phishing sites that focuses on the features of URL (PrimaryDomain, SubDomain, PathDomain) and the ranking of site (PageRank, AlexaRank, AlexaReputation. Then, a proposed neuro-fuzzy network is a system which reduces the error and increases the performance. The proposed neuro-fuzzy model uses computational models to perform without rule sets. The proposed solution achieved identification accuracy above 99% with low false signals.

The rest of this paper is organized as follows: Section II presents the related works. System design is shown in section III. Section IV evaluates the accuracy of the method. Finally, Section V concludes the paper and figures out the future works.

II. RELATED WORK

The phishing identification techniques are classified into three categories such as blacklist, heuristic and machine learning. In the first approach, the phishing identification technique [2][3][4][5] maintains a list of phishing websites called blacklist. The blacklist technique is inefficient due to the rapid growth in the number of phishing sites. Therefore, the heuristic and machine learning approaches have received more attraction of researchers.

Cantina [6] presented the algorithm TF-IDF based on 27 features of webpage. This technique can identify 97% phishing sites with 6% false positives. Although this technique is efficient, the time extracting 27 features of webpage is too long to meet real time demand and some features are not necessary for improving the phishing identification accuracy. Similarly, Cantina+ [7] used machine learning techniques based on 15 features of webpage and only six of 15 features are efficient for phishing identification such as bad form, Bad action fields, Non-matching URLs, Page in top search results, Search copyright brand plus domain and Search copyright brand plus hostname. In [8], the author used the URL to identify phishing sites automatically by extracting and verifying different terms of a URL through search engine. Even though this paper proposed a new interesting technique, the identification rate is quite low (54.3%). The technique [9] developed a content-based approach to identify phishing called CANTINA, which considers the Google PageRank value of a page, the evaluation dataset is quite small. The characteristic of the source code is used to identify phishing sites in [10].

The authors in [11] have proposed fuzzy technique based on 27 features of webpage, classified into 3 layer. Each feature has three linguistic values: low, moderate, high. The technique has built a rule set, triangular and trapezoidal membership functions. The achieved rate of the technique is 86.2%. But, there exist many drawbacks in [11]. First, the rule sets are not objective and greatly depend on the builder.

This research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)” (NIPA-2014-(H0301-14-1003).

L.A.T Nguyen, B.L To and H.K Nguyen are with Faculty of Information Technology, Ho Chi Minh City University of Transport, Vietnam nlatuan@hcmutrans.edu.vn, tblam83@gmail.com, nhkhuong@hcmutrans.edu.vn

C. Pham and C.S Hong are with Kyung Hee University, Korea pchuan@khu.ac.kr, cshong@khu.ac.kr

Second, the weight of each main criteria is used without any clarification. Finally, the used heuristics are not optimal and really effective.

The authors in [12] have proposed neural network technique. Three layers were used in the neural network including input layer, hidden layer and output layer. The best achieved rate of the technique is 95%. However, there exist some drawbacks in [12]. First, a number of hidden nodes and activation function must be determined through experimentation. Second, the authors do not explain why using one hidden layer. Third, the value of features do not know how is it calculated. Finally, the datasets are not big enough.

In the previous techniques, the URL plays a minor role in identifying phishing websites. In this paper, we focus on URL's features and design a new neuro-fuzzy model to identify phishing sites. Our work contributes four new aspects: i) The new heuristics have been proposed to identify phishing website more effectively and rapidly. ii) The parameter values used in the membership functions are derived from the big data set so that the model is still equivalent for the new data set. iii) The weights are trained by neural network, so they were more efficient. iv) The rule sets are not utilized. Hence, the result will be more precise and objective.

III. SYSTEM DESIGN

A. URL

A URL (uniform resource locator) is used to locate the resources[13].

The structure of URL is as follows:

$< protocol >: // < subdomain > . < primarydomain > . < TLD > / < pathdomain >$

For example, with the URL: http://www.paypal.abc.net/login/web/index.html, there are six components as follows: Protocol is http, Subdomain is paypal, Primarydomain is abc, TLD is net, Domain is abc.net, Pathdomain is login/web/index.html

B. Features of URL

Phishers usually try to make the Internet address (URL) of phishing sites look similar to legitimate sites to fool online users. They cannot use the exact URL of the legitimate site, they make more spelling mistake the features of URL such as PrimaryDomain, SubDomain, PathDomain. For example, the URL www.applle.com looks similar to well known website www.apple.com, or http://www.apple.attack.com if users are not careful, they will think that they are on the Apple site.

C. Features of Domains Ranking

Obviously, the phishing sites are accessed by the users or linked by the other websites. Therefore, the ranking of site such as PageRank, AlexaRank, AlexaReputation can also help to identify phishing sites. Phishers usually make fake-site of famous site, but the ranking of fake-site is not high. We can also use the rankings to identify whether a site is a phishing site.

D. System Model Design

The model can be depicted in Fig 1.

1) Phase I - Selecting four features of URL: Four features are extracted from URL such as *Domain*, *PrimaryDomain*, *SubDomain* and *PathDomain*.

2) Phase II - Calculating six values of the heuristics: Six values of the heuristics are calculated and six heuristics are six input nodes of the neuro-fuzzy network.

3) Phase III - Neuro-Fuzzy Network: The neuro-fuzzy network performs to calculate the value of the output node.

4) Phase IV - Identifying the sites: We based on the value of the output node to decide whether a site is a phishing site.

E. Neuro-Fuzzy Network Model

1) The model: The neuro-fuzzy network model was designed as in Fig 2. The model was designed with five layers as follows:

- The first layer, called the input layer, contains six nodes that are six heuristics such as PrimaryDomain, SubDomain, PathDomain, PageRank, AlexaRank, AlexaReputation.
- The second layer contains 12 nodes. The value of each node is calculated from the left sigmoid membership functions and the right sigmoid membership function.
- The third layer contains two nodes which are π_L and π_P. π_L and π_P are calculated by (1) and (2).

$$\pi_L = \prod_{i=1}^{6} L_i \tag{1}$$

$$\pi_P = \prod_{i=1}^{6} P_i \tag{2}$$

- The fourth layer contains two nodes which are NL (Normalization Legitimate) and NP (Normalization Phishing). NL and NP are calculated by (3) and (4).

$$NL = \frac{\pi_L}{\pi_L + \pi_P} \tag{3}$$

$$NP = \frac{\pi_P}{\pi_L + \pi_P} \tag{4}$$

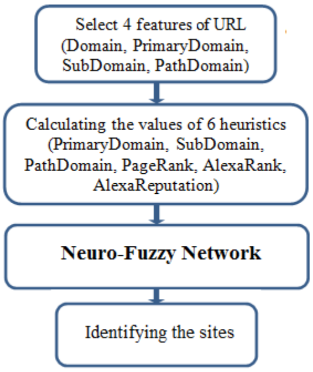

Select 4 features of URL (Domain, PrimaryDomain, SubDomain, PathDomain)
Calculating the values of 6 heuristics (PrimaryDomain, SubDomain, PathDomain, PageRank, AlexaRank, AlexaReputation)
Neuro-Fuzzy Network
Identifying the sites

Fig. 1. The System Model

- The fifth layer, called the output layer, has one output node.

The neural network performs from the fourth layer to the output layer. The weights are trained by the training algorithm and the sigmoid activation function is used in the proposed model, so the output value of the output node ranges from 0 to 1. The proposed model is classified into two classes so the site is phishing if the value of the output node is less than 0.5 and the site is legitimate, if the value is greater than or equal to 0.5.

2) The value of six input nodes: Based on experimental results and statistics from the dataset of 11,660 phishing sites. We found that:

- The phising site has the Levenshtein distance [14] between PrimaryDomain, SubDomain, PathDomain and the result of GOOGLE search engine spelling suggestion that is less than 4.
- The PageRank value varies from 0 to 10. The phishing site has the PageRank value that is less than 6.
- The phishing site has the AlexaRank value that is greater than 300,000.
- The phishing site has the AlexaReputation value that is less than 20.

Six values of the heuristics are calculated as follows:

- Calculating the value of heuristic PrimaryDomain: The algorithm is shown in Algorithm 1.
- Calculating the value of heuristic SubDomain and PathDomain: The algorithm is shown in Algorithm 2.
- Calculating the value of heuristic PageRank: The Googles PageRank value can be obtained from [15]. PageRank value varies from 0 to 10.
- Calculating the value of heuristic AlexaRank and AlexaReputation: AlexaRank and AlexaReputation value can be obtained from [16].

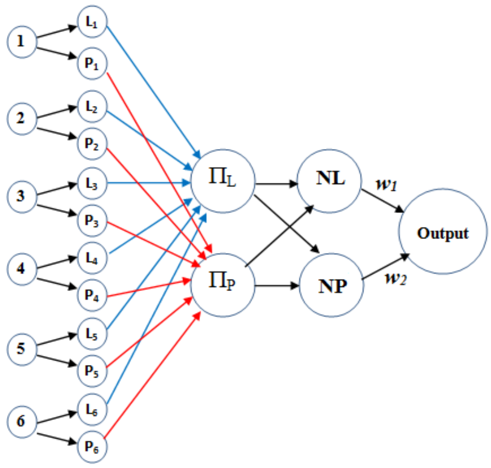


Fig. 2. The neuro-fuzzy network model

```
Data: PrimaryDomain
Result: The value of heuristic "PrimaryDomain"
if PrimaryDomain is IP then
    $value = 0$; //doubt phishing
else
    $Result = Suggestion\_Google(PrimaryDomain)$;
    if Result is NULL then
        $value = 100$; //No doubt phishing
    else
        $value = Levenshtein(Result, PrimaryDomain)$;
    end
end
```

Algorithm 1: Calculating the value of PrimaryDomain

```
Data: m //m is SubDomain or PathDomain
Result: The value of heuristic m
if m is Null then
    $value = 100$; //No doubt phishing
else
    $Result = Suggestion\_Google(m)$;
    if Result is NULL then
        $value = 100$; //No doubt phishing
    else
        $value = Levenshtein(Result, m)$;
    end
end
```

Algorithm 2: Calculating the value of SubDomain/PathDomain

3) The value of 12 nodes in the second layer: Classifying heuristics into two linguistic labels and assigning membership functions such as left sigmoid and right sigmoid for each of the linguistic value. Each of these heuristics is classified into linguistic labels as "Phishing" and "Legitimate". Based on experimental results and statistics from the dataset of 11,660 phishing sites, membership functions are calculated as follows:

- Membership functions for PrimaryDomain, SubDomain, PathDomain, "Pagerank" and "AlexaReputation": Equation (5) and (6) are two membership functions that are built to calculate fuzzy values and the graph of the membership functions is shown in Fig 3 .

$$L(x) = \frac{1}{1 + e^{-(x-b)}} \tag{5}$$

$$P(x) = \frac{e^{-(x-b)}}{1 + e^{-(x-b)}} \tag{6}$$

  Where parameter b for "PrimaryDomain", "SubDomain", "PathDomain", "Pagerank" and "AlexaReputation" are 4, 4, 4, 6 and 20, respectively.
- Membership functions for AlexaRank: Equation (7) and (8) are 2 membership functions built to calculate fuzzy values with parameter b of 300.000 and the graph of the membership functions is shown in Fig 4.

$$P(x) = \frac{1}{1 + e^{-(x-b)}} \tag{7}$$

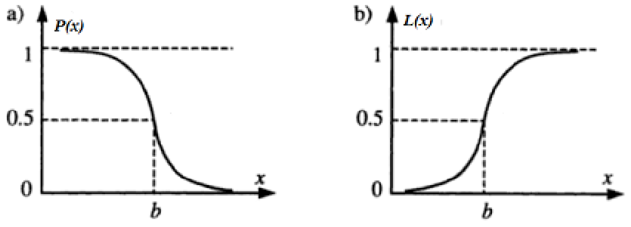


Fig. 3. Graph of membership function

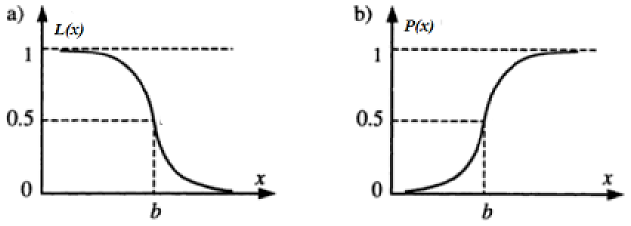


Fig. 4. Graph of membership function for ”AlexaRank”

$$L(x) = \frac{e^{-(x-b)}}{1 + e^{-(x-b)}} \tag{8}$$

4) Neural Network Training Algorithm: The proposed algorithm is shown in Fig 5. The algorithm performs two phases as follows:

- The *”propagation”* phase calculates the input value of the output node and the output value of the output node. The input value of the output node is calculated by (9)

$$O_I = \sum_{i=1}^{6} W_i * I_i \tag{9}$$

  Where O_I, I_i and W_i are the input value of the output node, the value of the ith input node and the weight of the ith input node respectively.
  The output value of the output node is calculated by (10)

$$O_O = \frac{1}{1 + e^{-O_I}} \tag{10}$$

  Where O_O and O_I are the output value of output node and the input value of output node respectively.
- The *”weight update”* phase calculates the error of the output node and updates the weights. The error of the output node is calculated by (11)

$$Err = O_O * (1 - O_O) * (T - O_O) \tag{11}$$

  Where T is the real value of sample in training dataset. The weights are updated by (12)

$$W_i = W_i + R * Err * O_O \tag{12}$$

  Where R and W_i are learning rate and the weight of the ith input node respectively.

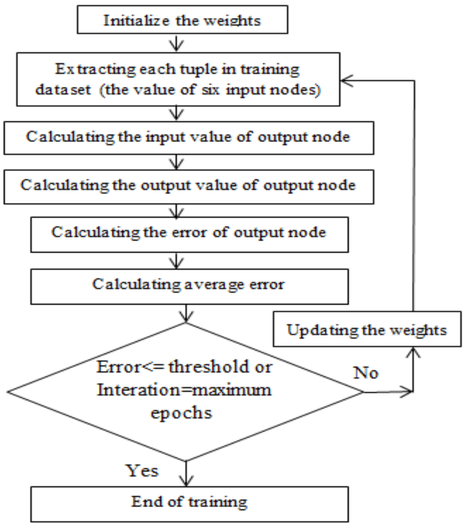

Initialize the weights
Extracting each tuple in training dataset (the value of six input nodes)
Calculating the input value of output node
Calculating the output value of output node
Calculating the error of output node
Calculating average error
Updating the weights
Error<= threshold or Interation=maximum epochs
No
Yes
End of training

Fig. 5. Neural Network Training Algorithm

IV. EVALUATION

We have collected 11,660 phishing sites from PhishTank [2] and 10,000 legitimate sites from DMOZ [17]. The training dataset contains 6,660 phishing sites from PhishTank and 5,000 legitimate sites from DMOZ. We build 2 testing datasets, each of which contains 5,000 phishing sites or 5,000 legitimate sites. Experimental procedure is divided into 2 phases (Training and Testing) through PHP and MYSQL.

A. Training Phase

1) Import Training Dataset: Training dataset is imported into MYSQL. The result is shown in the Fig 6.

2) Extracting four features of URL: Four features (PrimaryDomain, SubDomain, PathDomain and Domain) are extracted. Fig 7 shows the obtained result.

3) Calculating the value of six input nodes: Google search engine spelling suggestions and alexa.com are used to calculate the value of the input nodes. The result is shown in the Fig 8.

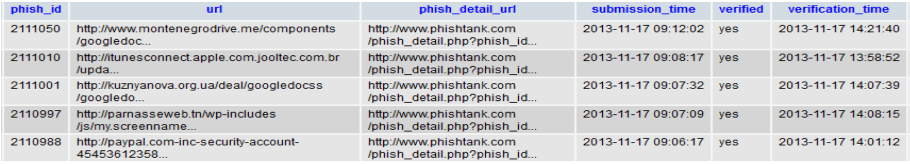
4) Calculating the fuzzy value of 12 nodes in the second layer : Two membership functions left sigmoid and right sigmoid are used to calculate the value of the nodes in the second layer. The result is shown in the Fig 9

5) Network Training phase: We performed the network training with 9 values of learning rate. In the training phase, the parameters are set as follows:

- Learning rate: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9
- Mean error threshold value: 1%
- Number of Epochs: 10,000
- The weights: initialize weights random values from 0 to 1

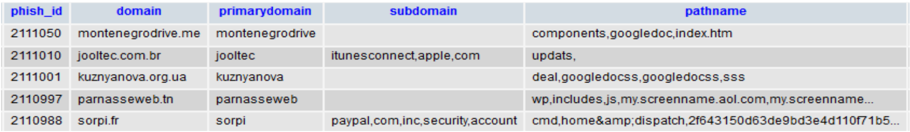
B. Testing Phase

In this phase, the proposed technique is tested with 2 testing datasets based on the weights of the network training



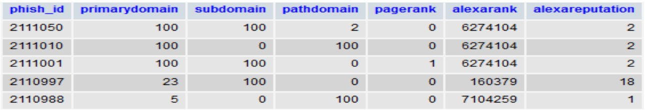
phish_id	url	phish_detail_url	submission_time	verified	verification_time
2111050	http://www.montenegrodrive.me/components /googledoc...	http://www.phishtank.com /phish_detail.php?phish_id...	2013-11-17 09:12:02	yes	2013-11-17 14:21:40
2111010	http://itunesconnect.apple.com.jooltec.com.br /upda...	http://www.phishtank.com /phish_detail.php?phish_id...	2013-11-17 09:08:17	yes	2013-11-17 13:58:52
2111001	http://kuznyanova.org.ua/deal/googledocss /googledo...	http://www.phishtank.com /phish_detail.php?phish_id...	2013-11-17 09:07:32	yes	2013-11-17 14:07:39
2110997	http://parnasseweb.tn/wp-includes /js/my.screenname...	http://www.phishtank.com /phish_detail.php?phish_id...	2013-11-17 09:07:09	yes	2013-11-17 14:08:15
2110988	http://paypal.com-inc-security-account-45453612358...	http://www.phishtank.com /phish_detail.php?phish_id...	2013-11-17 09:06:17	yes	2013-11-17 14:01:12

Fig. 6. Training dataset of 11,660 sites in MYSQL



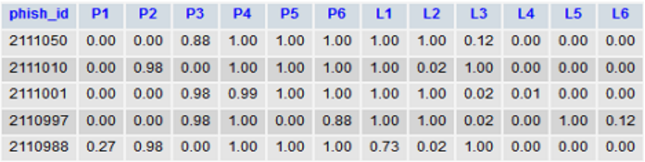
phish_id	domain	primarydomain	subdomain	pathname
2111050	montenegrodrive.me	montenegrodrive		components,googledoc,index.htm
2111010	jooltec.com.br	jooltec	itunesconnect,apple,com	updats,
2111001	kuznyanova.org.ua	kuznyanova		deal,googledocss,googledocss,sss
2110997	parnasseweb.tn	parnasseweb		wp,includes,js,my.screenname.aol.com,my.screenname...
2110988	sorpi.fr	sorpi	paypal,com,inc,security,account	cmd,home&amp;dispatch,2f643150d63de9bd3e4d110f71b5...

Fig. 7. Four features are extracted



phish_id	primarydomain	subdomain	pathdomain	pagerank	alexarank	alexareputation
2111050	100	100	2	0	6274104	2
2111010	100	0	100	0	6274104	2
2111001	100	100	0	1	6274104	2
2110997	23	100	0	0	160379	18
2110988	5	0	100	0	7104259	1

Fig. 8. Value of heuristics



phish_id	P1	P2	P3	P4	P5	P6	L1	L2	L3	L4	L5	L6
2111050	0.00	0.00	0.88	1.00	1.00	1.00	1.00	1.00	0.12	0.00	0.00	0.00
2111010	0.00	0.98	0.00	1.00	1.00	1.00	1.00	0.02	1.00	0.00	0.00	0.00
2111001	0.00	0.00	0.98	0.99	1.00	1.00	1.00	1.00	0.02	0.01	0.00	0.00
2110997	0.00	0.00	0.98	1.00	0.00	0.88	1.00	1.00	0.02	0.00	1.00	0.12
2110988	0.27	0.98	0.00	1.00	1.00	1.00	0.73	0.02	1.00	0.00	0.00	0.00

Fig. 9. Fuzzy values in the second layer

with learning rate of 0.1, 0.2, 0.3, 0.4, 0.5, 0.5, 0.6, 0.7, 0.8, 0.9. RMSE (Root Mean Square Error) is a good measure of identifying accuracy. RMSE is calculated by (13)

$$RMSE = \sqrt{\frac{\sum(A_i - I_i)^2}{N}} \tag{13}$$

Where I_i is the number of identifying sites, A_i is the number of actual sites and N is the number of samples in the testing dataset. Accuracy ratio is calculated as follows: Accuracy_Ratio = 100 - RMSE. The results of the test with learning rate of 0.1, 0.2, 0.3, 0.4, 0.5, 0.5, 0.6, 0.7, 0.8, 0.9 will be shown in Table I. From the obtained results, RMSE and accuracy are shown in Table II. We have found It shows the best ratio of 99.22% with learning rate of 0.7 and the worst ratio of 98.25% with learning rate of 0.2.

C. Comparing to technique [11]

We experimented with the technique [11] and compared to the result of our proposed technique. First, we collect 10 testing datasets, each of which contains 1,000 phishing sites or 1,000 legitimate sites. Second, we experiment the technique [11] and the results will be shown in Table III. From the obtained result and using RMSE, we have found that the technique [11] with the accuracy of 86.06%.

D. Comparing to technique [12]

We experimented with the technique [12] using 8 hidden nodes and hyperbolic tangent activation function. First, we collect 2 testing datasets, each of which contains 5,000 phishing sites or 5,000 legitimate sites. Second, we experiment the technique [12] and the results will be shown in Table IV. Then, the obtained results of RMSE and accuracy are shown in Table V. By using the technique in [12], we obtained the best accuracy of 94.68%.

V. CONCLUSIONS AND FUTURE WORKS

We have proposed a new technique to identify phishing sites effectively. In the technique, the system model is built to identify phishing sites by using neuro-fuzzy network and six heuristics (primarydomain, subdomain, pathdomain, pagerank, alexarank, alexareputation). The technique is experimented with the training dataset containing 11,660 sites and 2 testing datasets that each dataset contains 5,000 phishing sites or 5,000 legitimate sites. The best results show that 99.22% phishing websites are identified by using the system model. Our work is compared to the results in [11], [12] and found that it is more efficient. In the future, our neuro-fuzzy model will be improved to enhance the identification ratio. Besides, the system could be furthermore enhanced by using larger datasets and more heuristic parameters.

References

[1] Ollman, G. (2004) The Phishing Guide —Understanding and Preventing. White Paper, Next Generation Security Software Ltd.

[2] PhishTank. (2013, Nov.) Statistics about phishing activity and phishtank usage. [Online]. Available: http://www.phishtank.com/stats/2013/01/

[3] D. Goodin. (2012) Google bots detect 9,500 new malicious websites every day. [Online]. Available: http://arstechnica.com/security/2012/06/

[4] S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang. (2009) An empirical analysis of phishing blacklists. [Online]. Available: http://ceas.cc/2009/papers/ceas2009-paper-32.pdf

[5] McAfee. (2011, July) Mcafee site advisor. [Online]. Available: http://www.siteadvisor.com

[6] Y. Zhang, J. I. Hong, and L. F. Cranor, Cantina: a content-based approach to detecting phishing web sites, in The 16th international conference on World Wide Web, 2007, pp. 639—648

[7] G. Xiang, J. Hong, C. P. Rose, and L. Cranor, Cantina+: a feature-rich machine learning framework for detecting phishing web sites, ACM Transactions on Information and System Security, vol.14, no.2 .pp. 1—28, Sept. 2011.

[8] M. E. Maurer and D. Herzner, Using visual website similarity for phishing detection and reporting, in CHI 12 Extended Abstracts on Human Factors in Computing Systems, 2012, pp. 1625—1630.

[9] A. Sunil and A. Sardana, A pagerank based detection technique for phishing web sites, in IEEE Symposium on Computers & Informatics, 2012, pp. 58—63.

[10] M. G. Alkhozae and O. A. Batarfi, Phishing websites detected based on phishing characteristic in the webpage source code, in International Journal of Information and Communication Technology Research, vol. 1, no. 6, Oct. 2011, pp. 283—291

[11] M. Aburrous, M. Hossain, F. Thabatah, and K. Dahal, Intelligent phishing website detection system using fuzzy techniques, in Third International Conference on Information and Communication Technologies: From Theory to Applications, 2008, pp. 1—6.

[12] N. Zhang and Y. Yuan, Phishing Detection Using Neural Network, CS229 lecture notes, http://cs229.stanford.edu/proj2012/ZhangYuan-PhishingDetectionUsingNeuralNetwork.pdf, 2012

[13] Wikipedia. [Online]. Available (2014) : http://en.wikipedia.org/wiki/Uniformresourcelocator

[14] Levenshtein. [Online]. Available (2014) : http://en.wikipedia.org/wiki/Levenshteindistance

[15] G. Inc. [Online]. Available (2014) : http://toolbarqueries.google.com

[16] Alexa. [Online]. Available (2014) : http://data.alexa.com/data?cli=10&dat=snbamz&url=

[17] DMOZ. [Online]. Available (2014) : http://rdf.dmoz.org/rdf/

TABLE I

Result of testing with proposed technique

Learning Rate	Testing dataset	A_i	I_i
0.1	No.1	5,000	4,925
0.1	No.2	5,000	4,917
0.2	No.1	5,000	4,911
0.2	No.2	5,000	4,914
0.3	No.1	5,000	4,926
0.3	No.2	5,000	4,933
0.4	No.1	5,000	4,946
0.4	No.2	5,000	4,935
0.5	No.1	5,000	4,933
0.5	No.2	5,000	4,927
0.6	No.1	5,000	4,925
0.6	No.2	5,000	4,927
0.7	No.1	5,000	4,963
0.7	No.2	5,000	4,959
0.8	No.1	5,000	4,914
0.8	No.2	5,000	4,915
0.9	No.1	5,000	4,920
0.9	No.2	5,000	4,912

TABLE II

RMSE and Accuracy with proposed technique

Learing rate	RMSE	Accuracy
0.1	1.58	98.42%
0.2	1.75	98.25%
0.3	1.41	98.59%
0.4	1.20	98.80%
0.5	1.40	98.60%
0.6	1.48	98.52%
0.7	0.78	99.22%
0.8	1.71	98.29%
0.9	1.68	98.32%

TABLE III

Result of testing with technique [11]

(1):Very Phishy and Phishy (2) : Very Legitimate and Legitimate (3) : Suspicious

Testing dataset	(1)	(2)	(3)
No.1	867	82	51
No.2	865	76	59
No.3	847	90	63
No.4	902	172	26
No.5	841	109	50
No.6	64	873	63
No.7	50	911	39
No.8	39	895	66
No.9	97	871	32
No.10	85	863	52

TABLE IV

Result of testing with technique [12]

Learning Rate	Testing dataset	A_i	I_i
0.1	No.1	5,000	4,612
0.1	No.2	5,000	4,520
0.2	No.1	5,000	4,624
0.2	No.2	5,000	4,478
0.3	No.1	5,000	4,689
0.3	No.2	5,000	4,735
0.4	No.1	5,000	4,456
0.4	No.2	5,000	4,792
0.5	No.1	5,000	4,732
0.5	No.2	5,000	4,736
0.6	No.1	5,000	4,721
0.6	No.2	5,000	4,678
0.7	No.1	5,000	4,599
0.7	No.2	5,000	4,725
0.8	No.1	5,000	4,772
0.8	No.2	5,000	4,697
0.9	No.1	5,000	4,719
0.9	No.2	5,000	4,699

TABLE V

RMSE and Accuracy with technique [12]

Learing rate	RMSE	Accuracy
0.1	8.73	91.27%
0.2	9.10	90.90%
0.3	5.78	94.22%
0.4	8.24	91.76%
0.5	5.32	94.68%
0.6	6.03	93.97%
0.7	6.88	93.12%
0.8	5.36	94.64%
0.9	5.82	94.18%