Intelligent Phishing Website Detection System using Fuzzy Techniques

Maher Aburrous
Dept. of Computing
University of Bradford
Bradford, UK
mrmaburr@bradford.ac.uk

M. A. Hossain
Dept. of Computing
University of Bradford
Bradford, UK
m.a.hossain1@bradford.ac.uk

Fadi Thabatah
MIS Department
Philadelphia University
Amman, Jordan
ffayez@philadelphia.edu.jo

Keshav Dahal
Dept. of Computing,
University of Bradford
Bradford, UK
k.p.dahal@bradford.ac.uk

***Abstract*- Phishing websites are forged web pages that are created by malicious people to mimic web pages of real websites and it attempts to defraud people of their personal information. Detecting and identifying Phishing websites is really a complex and dynamic problem involving many factors and criteria, and because of the subjective considerations and the ambiguities involved in the detection, Fuzzy Logic model can be an effective tool in assessing and identifying phishing websites than any other traditional tool since it offers a more natural way of dealing with quality factors rather than exact values. In this paper, we present novel approach to overcome the 'fuzziness' in traditional website phishing risk assessment and propose an intelligent resilient and effective model for detecting phishing websites. The proposed model is based on FL operators which is used to characterize the website phishing factors and indicators as fuzzy variables and produces six measures and criteria's of website phishing attack dimensions with a layer structure. Our experimental results showed the significance and importance of the phishing website criteria (URL & Domain Identity) represented by layer one, and the variety influence of the phishing characteristic layers on the final phishing website rate.**

Keywords- Phishing; Fuzzy Logic; risk assessment; phishing website criteria

I. INTRODUCTION

PHISHING websites are forged web pages that are created by malicious people to mimic web pages of real websites. Most of these kinds of Web pages have high visual similarities to scam their victims. Some of these kinds of Web pages look exactly like the real ones. Unwary Internet users may be easily deceived by this kind of scam. Victims of phishing Web pages may expose their bank account, password, credit card number, or other important information to the phishing Web page owners. The impact is the breach of information security through the compromise of confidential data and the victims may finally suffer losses of money or other kinds. Phishing is a relatively new Internet crime in comparison with other forms, e.g., virus and hacking. More and more phishing Web pages have been found in recent years in an accelerative way [7]**.** The word phishing from the phrase "website phishing" is a variation on the word "fishing." The idea is that bait is thrown out with the hopes that a user will grab it and bite into it just like the fish. In most cases, bait is either an e-mail or an instant messaging site, which will take the user to hostile phishing websites [10].

Phishing website is a very complicate and complex issue to understand and to analyze, since it is joining technical and social problem with each other for which there is no known single silver bullet to solve it entirely. The motivation behind my study is to create a resilient and effective method that uses fuzzy logic to quantify and qualify all the website phishing characteristics and factors in order to detect phishing websites to assess whether phishing activity is taking place or not.

The paper is organized as follows: Section 2 presents the literature review and related work and Section 3 shows the theory and methodology of the website phishing risk assessment model with its system design and implementation. Section 4 reveals the experiments and results of the fuzzy phishing website risk assessment model and then conclusions and future work are given in Section 5.

II. LITERATURE REVIEW AND RELATED WORK

A. Literature Review

Phishing website is a recent problem, nevertheless due to its huge impact on the financial and on-line retailing sectors and since preventing such attacks is an important step towards defending against website phishing attacks, there are several promising approaches to this problem and a comprehensive collection of related works. In this section, we briefly survey existing anti-phishing solutions and list of the related works. One approach is to stop phishing at the email level [3], since most current phishing attacks use broadcast email (spam) to lure victims to a phishing website [21]. Another approach is to use security toolbars. The phishing filter in IE7 [19] is a toolbar approach with more features such as blocking the user's activity with a detected phishing site. A third approach is to visually differentiate the phishing sites from the spoofed legitimate sites. Dynamic Security Skins [5] proposes to use a randomly generated visual hash to customize the browser window or web form elements to indicate the successfully authenticated sites. A fourth approach is two-factor authentication, which ensures that the user not only knows a secret but also presents a security token [6]. However, this approach is a server-side solution. Phishing can still happen at sites that do not support two-factor authentication. Sensitive information that is not related to a specific site, *e.g.*, credit

card information and SSN, cannot be protected by this approach either [22].

Many industrial antiphishing products use toolbars in Web browsers, but some researchers have shown that security tool bars don't effectively prevent phishing attacks. In [5] Rachna Dhamija and Doug Tygar proposed a scheme that uses a cryptographic identity-verification method that lets remote Web servers prove their identities. However, the proposal requires changes to the entire Web infrastructure (both servers and clients), so it can succeed only if the entire industry supports it. Reference [13] also proposed a tool to model and describes phishing by visualizing and quantifying a given site's threat, but this method still wouldn't provide an antiphishing solution. Another approach is using certification, e.g., (microsoft.com/mscorp/twc/privacy/spam), [14], [15], [17], [1]. A recent and particularly promising solution [8] proposes to combine the technique of standard certificates with a visual indication of correct certification; a site-dependent logo indicating that the certificate was valid, would be displayed in a trusted credentials area of the browser. A variant of web credential is to use a database or list published by a trusted party, where known phishing web sites are blacklisted. For example Netcraft antiphishing toolbar http://toolbar.netcraft.com/ prevents phishing attacks by using a centralized blacklist of current phishing URLs. Other Examples include Websense, McAfee's anti–phishing filter, Netcraft anti-phishing system, Cloudmark SafetyBar, Microsoft Phishing Filter [16]**.** The weakness of this approach is its poor scalability and its timeliness. Note that phishing sites are cheap and easy to build and their average lifetime is only a few days.

APWG provides a solution directory at (Anti-Phishing Working Group) [2] which contains most of the major antiphishing companies in the world. However, an automatic antiphishing method is seldom reported. The typical technologies of antiphishing from the User Interface aspect are done by [5] and [22]. They proposed methods that need Web page creators to follow certain rules to create Web pages, either by adding dynamic skin to Web pages or adding sensitive information location attributes to HTML code. However, it is difficult to convince all Web page creators to follow the rules [7]. In [12], [7], [13], the DOM-based [20] visual similarity of Web pages is oriented, and the concept of visual approach to phishing detection was first introduced. Through this approach, a phishing Web page can be detected and reported in an automatic way rather than involving too many human efforts. Their method first decomposes the Web pages (in HTML) into salient (visually distinguishable) block regions. The visual similarity between two Web pages is then evaluated in three metrics: block level similarity, layout similarity, and overall style similarity, which are based on the matching of the salient block regions [7].

B. Main Characteristics of phishing websites.

Evolving with the antiphishing techniques, various phishing techniques and more complicated and hard-to-detect methods are used by phishers. The most straightforward way for a phisher to defraud people is to make the phishing Web pages similar to their targets.

Actually, there are many characteristics and factors that can distinguish the original legitimate website from the forged faked phishing website like Spelling errors, Long URL address and Abnormal DNS record [16]. The full list is shown in table I which will be used later on our analysis and methodology study.

C. Why using Fuzzy Logic?

Fuzzy logic has been used for decades in the engineering sciences to embed expert input into computer models for a broad range of applications. It offers a promising alternative for measuring operational risks [18]. The fuzzy logic approach provides more information to help risk managers effectively manage assessing and ranking website phishing risks than the current qualitative approaches as the risks are quantified based on a combination of historical data and expert input. The advantage of the fuzzy approach is that it enables processing of vaguely defined variables, and variables whose relationships cannot be defined by mathematical relationships [23]. Fuzzy logic can incorporate expert human judgment to define those variable and their relationships.

III. THEORY & METHODOLOGY

A. The Phishing Website Risk Assessment Model

1) Fuzzification

The approach described here is to apply fuzzy logic modeling to assess website phishing risk on the 27 characteristics and factors which stamp the forged website. The essential advantage offered by fuzzy logic techniques is the use of ***linguistic variables*** to represent Key Phishing Characteristic Indicators and relating website phishing probability. In this step, linguistic descriptors such as High, Low, Medium, for example, are assigned to a range of values for each Key Phishing Characteristic Indicators. Valid ranges of the inputs are considered and divided into classes, or fuzzy sets. For example, length of URL address can range from 'low' to 'high' with other values in between. We cannot specify clear boundaries between classes. The degree of belongingness of the values of the variables to any selected class is called the degree of membership; Membership function is designed for each Phishing characteristic indicator, which is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between [0, 1]. Linguistic values are assigned for each Phishing indicator as *Low, Moderate, and High* while for Phishing website risk rate as *Very legitimate, Legitimate, Suspicious, Phishy, and Very phishy* (*triangular and trapezoidal membership function*). For each input their values ranges from 0 to 10 while for output, ranges from 0 to 100.

An example of the linguistic descriptors used to represent one of the Key Phishing Characteristic Indicators (*URL Address*

Long) and a plot of the ***fuzzy membership functions*** are shown in figure 1. The fuzzy representation more closely matches human cognition, thereby facilitating expert input and more reliably representing experts' understanding of underlying dynamics [4].

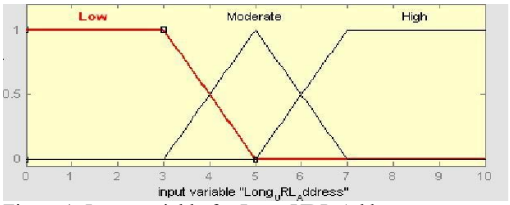

Figure 1. Input variable for Long URL Address component

The same approach is used to calibrate the other 26 Key Phishing Characteristic Indicators.

2) Rule Evaluation.

Having specified the risk of website phishing and its Key Phishing Characteristic Indicators, the logical next step is to specify how the website phishing probability varies as a function of the Key Phishing Characteristic Indicators. Experts provide fuzzy rules in the form of ***if…then*** statements that relate website phishing probability to various levels of Key Phishing Characteristic Indicators based on their knowledge and experience.

Website phishing experiments, Anti phishing tools analysis, web surveys, phishing quizzes and detailed questionnaire to assess factors, which collectively characterise the website phishing. A detailed checklist table is based on the types of phishing source and style, and weights assigned to them according to their effectiveness and influence.

3) Aggregation of the rule outputs.

This is the process of unification of the outputs of all rules. Combining the membership functions of all the rules consequents previously scaled into single fuzzy sets (output).

4) Defuzzification.

This is the process of transforming a fuzzy output of a fuzzy inference system into a crisp output. Fuzziness helps to evaluate the rules, but the final output this system has to be a crisp number. The input for the defuzzification process is the aggregate output fuzzy set and the output is a number. This step was done using Centroid technique because it is most commonly used method of defuzzification.

The output is website phishing risk rate and is defined in fuzzy sets like '**very phishy**' to '**very legitimate'**. The fuzzy output set is then defuzzified to arrive at a scalar value.

C. System Design

The design based on multi-level fuzzy approach for risk analysis [23]. Website phishing detection rate is performed based on six criteria: **URL & Domain Identity, Security & Encryption, Source Code & Java script, Page Style & Contents, Web Address Bar And Social Human Factor** as shown in Table I, which also shows that there are different number of components for each criterion, five components for **URL & Domain Identity, Source Code & Java script, Page Style & Contents, Web Address Bar,** four components for **Security & Encryption** and three components for **Social Human Factor.** Therefore, there are twenty seven components in total.

There are three layers on this website phishing fuzzy model as shown in figure 2. The first layer contains only **URL & Domain Identity** criteria with a weight equal to 0.3 for its importance; the second layer contains **Security & Encryption** criteria and **Source Code & Java script** criteria with a weight equal to 0.2 each; the third layer contains **Page Style & Contents** criteria**, Web Address Bar** criteria **And Social Human Factor** criteria with a weight equal to 0.1 each. The six criteria have been prioritized according to their importance using weights as concluded from the Website phishing experiments, case studies, Anti phishing tools analysis, web surveys, phishing quizzes, detailed questionnaire and phishing expert's feedback.

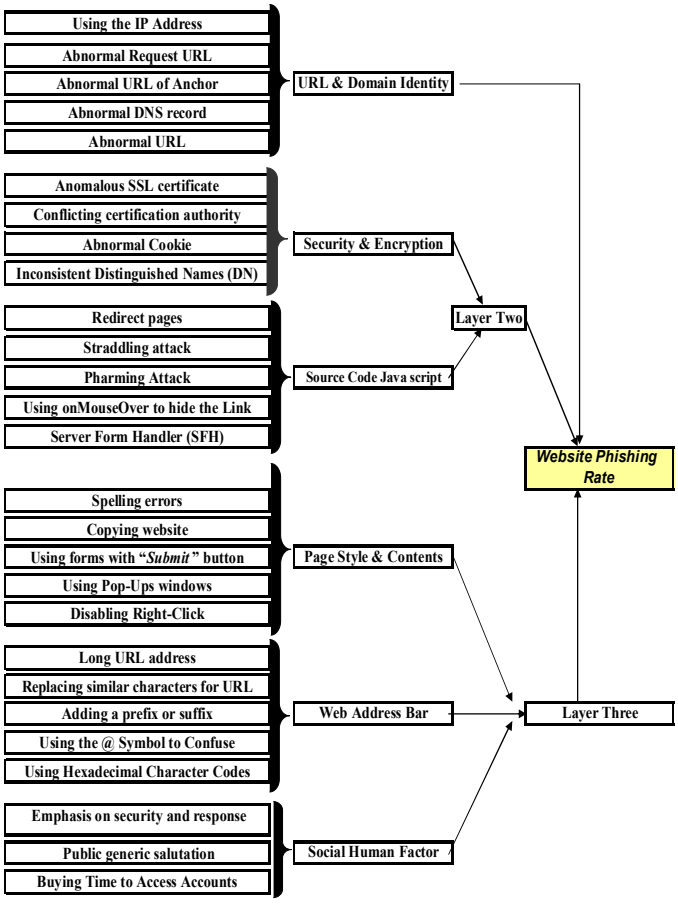

Figure 2. Structure of the fuzzy inference overall system to evaluate website phishing rate.

Website Phishing Rating = 0.3 * URL & Domain Identity crisp [First layer] + ((0.2 * Security & Encryption crisp)+(0.2 * Source Code & Java script crisp)) [Second layer] + ((0.1 * Page Style & Contents crisp) +(0.1 * Web Address Bar crisp) + (0.1 * Social Human Factor crisp)) [Third layer]

TABLE I. COMPONENTS AND LAYERS OF WEBSITE PHISHING CRITERIA.

Criteria	N	Component	Layer No.
URL & Domain Identity (Weight = 0.3)	1	Using the IP Address	Layer One; Sub weight = 0.3
	2	Abnormal Request URL	
	3	Abnormal URL of Anchor	
	4	Abnormal DNS record	
	5	Abnormal URL	
Security & Encryption (Weight = 0.2)	1	Using SSL certificate	Layer Two; Sub weight = 0.4
	2	Certification authority	
	3	Abnormal Cookie	
	4	Distinguished Names Certificate(DN)	
Source Code & Java script (Weight = 0.2)	1	Redirect pages	
	2	Straddling attack	
	3	Pharming Attack	
	4	Using onMouseOver to hide the Link	
	5	Server Form Handler (SFH)	
Page Style & Contents (Weight =0.1)	1	Spelling errors	Layer Three; Sub weight = 0.3
	2	Copying website	
	3	Using forms with "*Submit*" button	
	4	Using Pop-Ups windows	
	5	Disabling Right-Click	
Web Address Bar (Weight = 0.1)	1	Long URL address	
	2	Replacing similar characters for URL	
	3	Adding a prefix or suffix	
	4	Using the @ Symbol to Confuse	
	5	Using Hexadecimal Character Codes	
Social Human Factor (Weight = 0.1)	1	Much emphasis on security and response	
	2	Public generic salutation	
	3	Buying Time to Access Accounts	
Total Weight			**1**

D. The Rule Base

1) The Rule Base1 for layer 1.

The rule base has five input parameters and one output and contains all the "IF-THEN" rules of the system. For each entry of the rule base, each component is assumed to be one of three values and each criterion has five components.

TABLE II. SAMPLE OF THE RULE BASE 1 STRUCTURE AND ENTRIES FOR URL & DOMAIN IDENTITY CRITERIA

Rule #	(comp. 1) Using the IP Address	(comp. 2) Abnormal Req. URL	(comp. 3) Abnormal URL Anchor	(comp. 4) Abnormal DNS record	(comp. 5) Abnormal URL	URL & Domain Identity Criteria Phishing Risk *(Layer one)*
1	Low	Low	Low	Low	Low	Genuine
2	Low	Low	Low	Low	Mod.	Genuine
3	Low	Low	Low	Mod.	Mod.	Doubtful
4	Low	Low	Low	Mod.	high	Doubtful
5	Low	Low	Mod.	Mod.	high	Fraud
6	Low	Mod.	Mod.	Low	high	Fraud
7	Mod.	Low	high	Mod.	high	Fraud
8	high	Mod.	Low	Low	Low	Doubtful
9	Low	high	Low	Low	Mod.	Doubtful
10	high	Mod.	high	high	Low	Fraud

Therefore, the rule base 1 contains $(3^5) = 243$ entries. The output of rule base 1 is one of the website phishing rate fuzzy sets (Genuine, Doubtful or Fraud) representing URL & Domain Identity criteria phishing risk rate. A sample of the structure and the entries of the rule base 1 for layer 1 are shown in Table II. The system structure for URL & Domain Identity criteria is the joining of its five components (Using the IP Address, Abnormal Request URL, Abnormal URL of Anchor, Abnormal DNS record and Abnormal URL), which produces the URL & Domain Identity criteria (Layer one).

2) The Rule Base for layer 2.

In Layer 2, there are two inputs, which are (Security & Encryption and Source Code & Java script) and one output. The system structure for Security & Encryption criteria is the joining of its four components (Using SSL certificate, Certification authority, Abnormal Cookie and Distinguished Names Certificate(DN)) using Rule base 1, which produces Security & Encryption criteria. The system structure for Source Code & Java script criteria is the joining of its five components (Redirect pages, Straddling attack, Pharming Attack, Using onMouseOver to hide the Link and Server Form Handler (SFH)) using Rule base 1, which produces Source Code & Java script criteria. The structure and the entries of the rule base for layer 2 are illustrated in Table III. The system structure for layer 2 is the combination of two website phishing criteria (Security & Encryption and Source Code & Java script), which produces rule base 2. The rule base contains $(3^2) = 9$ entries and the output of rule base 2 is one of the website phishing rate fuzzy sets (Legal, Uncertain or Fake) representing Layer Two criteria phishing risk rate.

TABLE III. THE RULE BASE 2 STRUCTURE AND ENTRIES FOR LAYER TWO

Rule	Security & Encryption	Source Code & Java script	Phishing Risk (Layer Two)
1	Genuine	Genuine	Legal
2	Genuine	Doubtful	Legal
3	Genuine	Fraud	Uncertain
4	Doubtful	Genuine	Legal
5	Doubtful	Doubtful	Uncertain
6	Doubtful	Fraud	Uncertain
7	Fraud	Genuine	Uncertain
8	Fraud	Doubtful	Uncertain
9	Fraud	Fraud	Fake

3) The Rule Base for layer 3.

In Layer 3, there are three inputs, which are: the Page Style & Contents, Web Address Bar and Social Human Factor which is the output from layer 3, and one output. The system structure for Page Style & Contents criteria is the joining of its five components (Spelling errors, Copying website, Using forms with "*Submit*" button, Using Pop-Ups windows and Disabling Right-Click) using Rule base 1, which produces Page Style & Contents criteria. The system structure for Web Address Bar criteria is the joining of its five components (Long URL address, Replacing similar characters for URL, Adding a prefix or suffix, Using the @ Symbol to Confuse and Using Hexadecimal Character Codes) using Rule base 1,

which produces Web Address Bar criteria. The system structure for Social Human Factor criteria is the joining of its three components (Much emphasis on security and response, Public generic salutation and Buying Time to Access Accounts) using Rule base 1, which produces Social Human Factor criteria.

A sample of the structure and the entries of the rule base for layer 3 are shown in Table IV. The system structure for layer 3 is the combination of Page Style & Contents, Web Address Bar and Social Human Factor, which produces rule base 3. The rule base contains $(3^3) = 27$ entries and the output of rule base 3 is one of the website phishing rate fuzzy sets (Legal, Uncertain or Fake) representing Layer Three criteria phishing risk rate.

TABLE IV. THE RULE BASE 3 STRUCTURE AND ENTRIES FOR LAYER THREE

Rule	Page Style & Contents	Web Address Bar	Social Human Factor	Phishing Risk (Layer Three)
1	Genuine	Genuine	Genuine	Legal
2	Genuine	Doubtful	Fraud	Uncertain
3	Genuine	Fraud	Fraud	Fake
4	Doubtful	Genuine	Genuine	Legal
5	Doubtful	Doubtful	Doubtful	Uncertain
6	Doubtful	Fraud	Doubtful	Uncertain
7	Fraud	Genuine	Genuine	Legal
8	Fraud	Doubtful	Doubtful	Uncertain
9	Fraud	Fraud	Fraud	Fake

4) *The Rule Base for final website phishing rate.*

In the website phishing rule base last phase, there are three inputs, which are: layer one, layer two and layer three, and one output which is the rate of the phishing website. The structure and the entries of the rule base for website phishing rate are shown in Table V. The system structure for is the combination of layer one, layer two and layer three, which produces final website phishing rule base. The three dimensional plots of this structure is shown in Figure 3 using MATLAB. The rule base contains (33) = 27 entries and the output of final website phishing rule base is one of the final output fuzzy sets (Very Legitimate, Legitimate, Suspicious, Phishy or Very Phishy) representing final phishing website rate.

TABLE V. THE WEBSITE PHISHING RATE RULE BASE STRUCTURE AND ENTRIES FOR FINAL PHISHING RATE

Rule	URL & Domain Identity	Layer Two	Layer Three	Final Website Phishing Rate
1	Genuine	Legal	Legal	Very Legitimate
2	Genuine	Legal	Uncertain	Legitimate
3	Genuine	Legal	Fake	Suspicious
4	Genuine	Uncertain	Legal	Suspicious
5	Genuine	Uncertain	Uncertain	Phishy
6	Genuine	Uncertain	Fake	Phishy
7	Genuine	Fake	Legal	Suspicious
8	Genuine	Fake	Uncertain	Phishy
9	Genuine	Fake	Fake	Very Phishy
10	Doubtful	Legal	Legal	Legitimate
11	Doubtful	Legal	Uncertain	Suspicious
12	Doubtful	Legal	Fake	Phishy
13	Doubtful	Uncertain	Legal	Suspicious
14	Doubtful	Uncertain	Uncertain	Suspicious
15	Doubtful	Uncertain	Fake	Phishy
16	Doubtful	Fake	Legal	Phishy
17	Doubtful	Fake	Uncertain	Phishy
18	Doubtful	Fake	Fake	Very Phishy
19	Fraud	Legal	Legal	Suspicious
20	Fraud	Legal	Uncertain	Suspicious
21	Fraud	Legal	Fake	Phishy
22	Fraud	Uncertain	Legal	Suspicious
23	Fraud	Uncertain	Uncertain	Suspicious
24	Fraud	Uncertain	Fake	Phishy
25	Fraud	Fake	Legal	Phishy
26	Fraud	Fake	Uncertain	Very Phishy
27	Fraud	Fake	Fake	Very Phishy

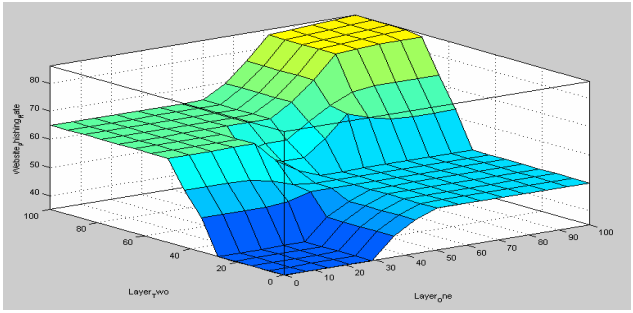

Figure 3. Three-dimensional plots for *final phishing rate*

IV. EXPERIMENTS AND RESULTS

Clipping method [9] is used in aggregating the consequences and the aggregated surface of the rule evaluation is defuzzified using Mamdani method [11] to find the Center Of Gravity (COG). Centroid defuzzification technique shown in equation (1) can be expressed as where x^* is the defuzzified output, $\mu i(x)$ is the aggregated membership function and x is the output variable.

$$x^* = \frac{\int \mu_i(x)\, x\, dx}{\int \mu_i(x)\, dx} \qquad \text{---------- (1)}$$

The proposed intelligent Phishing website detection system has been implemented in MATLAB 6.5. The results of some input combinations are listed in Tables VI, VII and VIII.

The final phishing website risk rating will be balanced (50%) representing a "***suspicious website***", when the Layer one (URL & Domain Identity) of the phishing website risk criteria has 10 input values which indicate *High* phishing indicator and all other layers have the value of zero inputs as shown in Table VI. Same result can be made when all phishing website risk criteria's representing by the three layers have middle (5) input values which indicate *Mod.* phishing indicator. These results shows the significance and importance of the phishing website criteria (URL & Domain Identity) represented by layer one especially when compared to the other criteria's and layers. Table VII shows that when the Layer one and Layer two of the phishing website risk criteria has middle (5) input values which indicate *Mod.* phishing indicator and other third Layer has the value of 10 input values which indicate *High* phishing indicator, the final phishing website risk rating will be reasonably high (65%) representing a "***phishy website***", which means that there is a Good guarantee that the website is forged phishy website. This result clearly shows that even if some of the website phishing characteristics or layers are not very clear or not definite, the website can still be phishy and forged and users should be aware when dealing with it especially when other phishing characteristics or layers are obvious and clear.

TABLE VI. FIVE HIGHEST (10) FOR LAYER ONE AND ALL OTHERS LOWEST (0).

Comp	Layer One URL & Domain	Layer Two		Layer Three			% Website Phishing Rating
		Security & Encrypt	Source Code & Java	Page Style & Contents	Web Address Bar	Social Human Factor	
1	10	0	0	0	0	0	**50%**
2	10	0	0	0	0	0	
3	10	0	0	0	0	0	
4	10	0	0	0	0		
5	10		0	0	0		

TABLE VII. FIVE MIDDLE (5) INPUTS FOR LAYER ONE AND LAYER TWO AND HIGHEST (10) INPUTS FOR LAYER THREE.

Comp	Layer One URL & Domain	Layer Two		Layer Three			% Website Phishing Rating
		Security & Encrypt	Source Code & Java	Page Style & Contents	Web Address Bar	Social Human Factor	
1	5	5	5	10	10	10	**65%**
2	5	5	5	10	10	10	
3	5	5	5	10	10	10	
4	5	5	5	10	10		
5	5		5	10	10		

TABLE VIII. FIVE MIDDLE (5) INPUTS FOR LAYER ONE AND ALL OTHERS LOWEST (0) INPUTS.

Comp	Layer One URL & Domain	Layer Two		Layer Three			% Website Phishing Rating
		Security & Encrypt	Source Code & Java	Page Style & Contents	Web Address Bar	Social Human Factor	
1	5	0	0	0	0	0	**35%**
2	5	0	0	0	0	0	
3	5	0	0	0	0	0	
4	5	0	0	0	0		
5	5		0	0	0		

Table VIII show that when the Layer one of the phishing website risk criteria (URL & Domain Identity) has middle (5) input values which indicate *Mod.* phishing indicator and all other Layers has the value of zero input values which indicate *Low* phishing indicator, the final phishing website risk rating will be reasonably low (35%) representing a "***legitimate website***", which means that there is Good guarantee that the website is legitimate website. This result clearly shows that even if some of the website phishing characteristics or layers are noticed or observed, that does not mean at all that the website is phishy or forged, but it can be safe and secured especially when other phishing characteristics or layers are not noticeable, visible or detectable. The results also indicates that the worst website phishing rate (all three layers have 10 input value) equals 86.2% representing "Very Phishy Website" and the best website phishing rate (all three layers have 0 input value) is 13.8% representing "Very Legitimate Website" rather than a full range, i.e. 0 to 100, because of the fuzzification process

V. CONCLUSION AND FUTURE WORK

The fuzzy website phishing model showed the significance and importance of the phishing website criteria (URL & Domain Identity) represented by layer one, and also showed that even if some of the website phishing characteristics or layers are not very clear or not definite, the website can still be phishy especially when other phishing characteristics or layers are obvious and clear. On the other hand even if some of the website phishing characteristics or layers are noticed or observed, that does not mean at all that the website is phishy, but it can be safe and secured especially when other phishing characteristics or layers are not noticeable, visible or detectable. As a future work we will propose and develop a prototype intelligent website phishing detection system by using Fuzzy Data Mining algorithms and techniques. The approach will look for deviations from stored patterns of normal phishing behavior and for previously described patterns of behavior that is likely to indicate phishing.

REFERENCES

[1] WholeSecurity Web Caller-ID, www.wholesecurity.com.

[2] Anti-Phishing Working Group. Phishing Activity Trends Report, http://antiphishing.org/reports/apwg_report_DEC2005_FINAL.pdf, December 2005.

[3] B. Adida, S. Hohenberger and R. Rivest, "Lightweight Encryption for Email," USENIX Steps to Reducing Unwanted Traffic on the Internet Workshop (SRUTI), 2005.

[4] S. M. Bridges and R. B. Vaughn, "fuzzy data mining and genetic algorithms applied to intrusion detection," Department of Computer Science Mississippi State University, White Paper, 2001.

[5] R. Dhamija and J.D. Tygar, "The Battle against Phishing: Dynamic Security Skins," Proc. Symp. Usable Privacy and Security, 2005.

[6] FDIC., "Putting an End to Account-Hijacking Identity Theft," http://www.fdic.gov/consumers/consumer/identity_theft.pdf, 2004.

[7] A. Y. Fu, L. Wenyin and X. Deng, " Detecting Phishing Web Pages with Visual Similarity Assessment Based on Earth Mover's Distance ," IEEE transactions on dependable and secure computing, pp. 301-311 vol. 3, no. 4, 2006.

[8] A. Herzberg and A. Gbara, "Protecting Naive Web Users," Draft of July 18, 2004.

[9] C. Y. Ho, B. W. Ling and J. D. Reiss, "Fuzzy Impulsive Control of High-Order Interpolative Low-Pass Sigma–Delta Modulators," IEEE Transactions on Circuits and Systems—I: Regular Papers, pp. 2224 - 2233 Vol. 53, No. 10, October 2006.

[10] L. James**,** "Phishing Exposed," Tech Target Article sponsored by: Sunbelt software, searchexchange.com, 2006.

[11] M. Liu, D. Chen and C. Wu. "The continuity of Mamdani method," International Conference on Machine Learning and Cybernetics, Page(s): 1680 - 1682 vol.3, 2002.

[12] W. Liu, G. Huang, X. Liu, M. Zhang, and X. Deng, "Phishing Web Page Detection," Proc. Eighth Int'l Conf. Documents Analysis and Recognition, pp. 560-564, 2005.

[13] W. Liu, X. Deng, G. Huang and A. Y. Fu, "An Antiphishing Strategy Based on Visual Similarity Assessment," Published by the IEEE Computer Society 1089-7801/06, INTERNET COMPUTING IEEE, pp. 58-65, 2006.

[14] Microsoft Corp, "Microsoft Phishing Filter: A New Approach to Building Trust in E-Commerce Content," White Paper, 2005.

[15] S. Olsen, "AOL tests caller ID for e-mail," CNET News.com, January 22, 2004.

[16] Y. Pan and X. Ding, "Anomaly BasedWeb Phishing Page Detection," Proceedings of the 22nd Annual Computer Security Applications Conference (ACSAC'06), Computer Society, pp. 381-392, 2006.

[17] J. C. Perez, "Yahoo airs antispam initiative," ComputerWeekly.com, December 8, 2003.

[18] S. Shah, "Measuring Operational Risks using Fuzzy Logic Modeling, "Article, Towers Perrin, JULY 2003.

[19] T. Sharif, "Phishing Filter in IE7," http://blogs.msdn.com/ie/archive/2005/09/09/463204.aspx, 2006.

[20] L. Wood, "Document Object Model Level 1 Specification," http://www.w3.org, 2005.

[21] M. Wu, R. C. Miller and S. L. Garfinkel , "*Do Security Toolbars Actually Prevent Phishing Attacks?,"* CHI April 2006.

[22] M. Wu, R. C. Miller and G. Little, "Web Wallet: Preventing Phishing Attacks by Revealing User Intentions," MIT Computer Science and Artificial Intelligence Lab, 2006.

[23] K.P. Dahal, Z. Hussain and M.A. Hossain," Loan risk analyzer based on fuzzy logic**,"** Proceedings of IEEE Int. Conference on E-Technology & E-Commerce, IEEE Comp. Society Press, pp 363-366.