A Proactive Anti-Phishing Tool Using Fuzzy Logic and RIPPER Data Mining Classification Algorithm

Rosana J. Ferolin

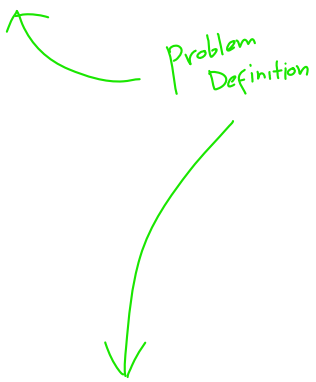
Department of Computer Engineering University of San Carlos, Cebu City, Philippines
rjferolin@usc.edu.ph

Abstract. Phishing emails are commonly used in fraud to illegally obtain confidential information mostly from clients of financial organizations. Phishing attacks are increasing in volume and have become more and more sophisticated which sometimes bypass the anti-phishing filter set by existing tools. This may be attributed to the fact that most of the approaches used by existing anti-phishing techniques are passive in form. Such approaches are passive in a way since they do not stop the source of the phishing emails or phishing websites but rather they simply classify and detect phishing emails.This paper presents a method for assessing and identifying phishing emails using Fuzzy Logic and the RIPPER Data Mining algorithm. In assessing the Phishing email, Fuzzy Logic linguistic descriptors are assigned to a range of values for each key phishing characteristic indicators. The membership function is designed for each Phishing email characteristic descriptor. The Data Mining RIPPER algorithm is used to characterize the Phishing emails and classify them based on both content-based and non-content based characteristics of Phishing emails. Furthermore, after the email has been assessed and classified as a Phishing email, the system proactively gets rid of the Phishing site or Phishing page by sending a notification to the Administrator of the host server that it is hosting a Phishing site which may result in the removal of the site.The initial results showed that the RIPPER algorithm achieved 85.4% for correctly classified Phishing emails and 14.6% for wrongly classified Phishing emails based on publicly available datasets from Phistank. After classifying the Phishing email, the system sends a notification to the System Administrator of the host server to indicate that it is hosting a Phishing site.

Keywords: Data Mining, Fuzzy Logic, Phishing, URL, Classification

1 Introduction

Phishing is as an act of sending an e-mail to a user falsely claiming to be a legitimate business establishment in an attempt to scam or trick the user into surrendering private information that will be used for identity theft. It is a type of



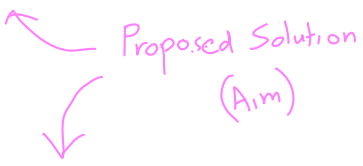
network attack where the attacker creates a replica of an existing legitimate commercial website to deceive users to submit personal, financial, or confidential data to what to they think is their genuine business provider's site. It is a security attack that involves obtaining private and classified data by presenting oneself as a reliable and genuine entity. The phishing e-mail then directs the user to visit a Web site that looks very similar to the Web site of the valid commercial establishment. Most of the time, Phishers take advantage of user's trust on the appearance of a site by designing websites that are very similar to an authentic establishment's website. The difference in appearance is most of the time unnoticeable by just merely looking at the site, where they are prompted to provide and update personal information such as passwords, bank account numbers, pin numbers, credit card numbers and other private information. This however, is a hoax and is set-up to steal the user's information. Once the private information is obtained, the hacker can now login to the user's account and process fraudulent transactions. This makes phishing an online fraud. The damage caused by phishing ranges from loss of access to email to significant and considerable financial loss.

There were at least 67, 677 phishing attacks reported by the Anti-Phishing Working Group (APWG) in the last six months 0f 2010 [1]. The latest reports showed that most phishing attacks are "spear phishing" that aim the financial, business and payment sectors [2]. The number of phishing attacks and phishing sites is rapidly increasing and on the average, Sophos identifies 16,173 malicious web pages everyday [3].

Even though Web browsers (i.e. Mozilla Firefox, Internet Explorer, Opera, Google Chrome, etc.) provide add-on tool for blocking phishing emails and phishing sites, Phishers still manage to override these security mechanisms. Phishing has become more and more complicated that Phishers can bypass the filter set by current anti- phishing techniques. The rapidly increasing number of phishing attacks suggests that it is therefore difficult to find a single logical procedure to detect phishing emails and that existing anti-phishing tools are not sufficient. This may be attributed to the mostly passive approach of anti-phishing techniques. The approaches are passive since they do not stop the source of the phishing emails rather they simply classify and detect phishing emails [4].

A proactive approach to minimizing phishing has been conducted where the system removes a phishing page from the host server rather than just filtering email and flagging suspected messages as spam [4]. The study in [4] however, assumes that emails have already been classified as a phishing email or legitimate email. The study has ignored the phishing email classification and was more concerned with how to deal with the Phisher once a phishing email has been detected.

This study proposes to develop an anti-phishing tool that combines phishing email classification and the proactive approach to stop the Phishers from its source as a result of the classification. This will be an improvement of the of the existing study in [4] where classification was ignored and that the proactive approach is based on the assumption that emails have been already identified as phishing emails. The study will take into consideration different email features in classifying phishing emails. The aim of this study is to use Fuzzy Logic and Data Mining Classification algorithm in classifying



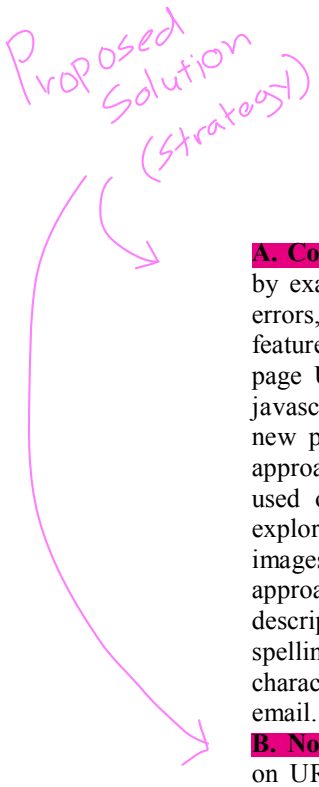
and detecting phishing emails. In addition, once an email is classified as a phishing email, the system will proactively stop phishing at its source by tracing the website and informing the host server of the of the phishing activity.

2 Related Studies

Most anti-phishing tools employ email filtering techniques to classify legitimate emails and suspected spam in the mail inbox. The user is left to decide whether to open or discard such emails. If no anti-phishing tool is installed or the user has not updated the anti-phishing program, then there is no layer of protection. This is referred to as passive anti-phishing [4]. It is because the approach only locally protects the user from a phishing attack but does not make any effort to stop or remove the Phisher at the source. The Phisher then continues with the phishing operation to further increase its victims.

While there are several email filters, browser tools, anti-spyware and anti-virus software, very few research efforts have been entirely focused to protect online users from phishing attacks in the past. Existing anti-phishing and anti-spam techniques suffer from one or more limitations and they are not 100% effective at stopping all spam and phishing attacks [9]. Phishers are able to find ways to bypass existing rule-based and statistical based filters without much difficulty. Major e-mail service providers such as Yahoo, Hotmail, Gmail, and AOL filter all incoming emails separating them into Inbox (legitimate email) and junk (illegitimate email) email folders. However, these e-mail service providers do not actually attempt to remove the phishing page associated with the illegitimate email. Furthermore, Phishers have readily available tools to bypass such spam filters [5]. We refer to this as a passive anti-phishing approach. This is because the approach only attempts to locally protect an individual from a phishing attack, but does not actively make any effort to remove or shut down the Phisher at the source. In effect, the Phisher is free to continue with the fraudulent operation and can potentially accrue further victims.

There are different existing phishing detection approaches. These approaches can be further classified as 1) content-based approaches that use site contents to detect phishing, 2) non-content based approaches that do not use the content of the site to identify if the email is an authentic or phishing email and 3) Visual based approaches that identify phishing using the similarity of known sites through visual inspection [5].



A. Content Based Approach. In content based approach, phishing attacks are detected by examining site contents. Features used in this approach include keywords, spelling errors, links, password fields, embedded links, etc. along with URL and host based features [5]. Google's anti-phishing filter detects phishing and malware by examining page URL, page rank, WHOIS information and contents of a page including HTML, javascript, images, iframe, etc. [5]. The classifier is constantly updated to accommodate new phishing sites to cope up with the latest techniques in phishing attacks. In this approach the classifier may have higher accuracy but the result is not real-time. It is used offline since it takes longer to detect the Phishing. Several researchers have explored different approaches such as fingerprinting, principal component analysis of images, heuristic approaches and fuzzy logic among others and fuzzy logic based approaches to identify phishing sites. Our approach uses Fuzzy Logic language descriptors with a range of values for each identified phishing characteristic specifically spelling errors, keywords and embedded links. The membership function for each characteristic derived as is used to assess the probability that the email is a phishing email.

B. Non-Content Based Approach. Non-content based approaches are primarily based on URL and host information classification. URLs are commonly classified based on features such as URL address length and presence of special characters. Moreover, host features of URL such as IP address, site owner, DNS properties and geographical properties are also used in the classification of Phishing emails [5]. The success rate is between 95% - 99% even in real-time processing [15].

3 Methodology

3.1 System Flow

This section describes the overall approach of the system in assessing, detecting and classifying Phishing emails. The process for notifying the hosting site or the sending site of the Phishing email is also included as we ll as the possible removal process. At the start, the system assesses the risk of the email using Fuzzy Logic. It then classifies the email as Phishing or legitimate email. The classification makes use of the data mining RIPPER algorithm. If the system detects that it is a phishing email, it gets the URL of the Phishing email. The host server's IP address, host server location and the contact information of the System Administrator. A notification is sent to the System Administrator of the host server informing that a phishing page is hosted by the server. The System Administrator proceeds with the removal of the Phishing page.

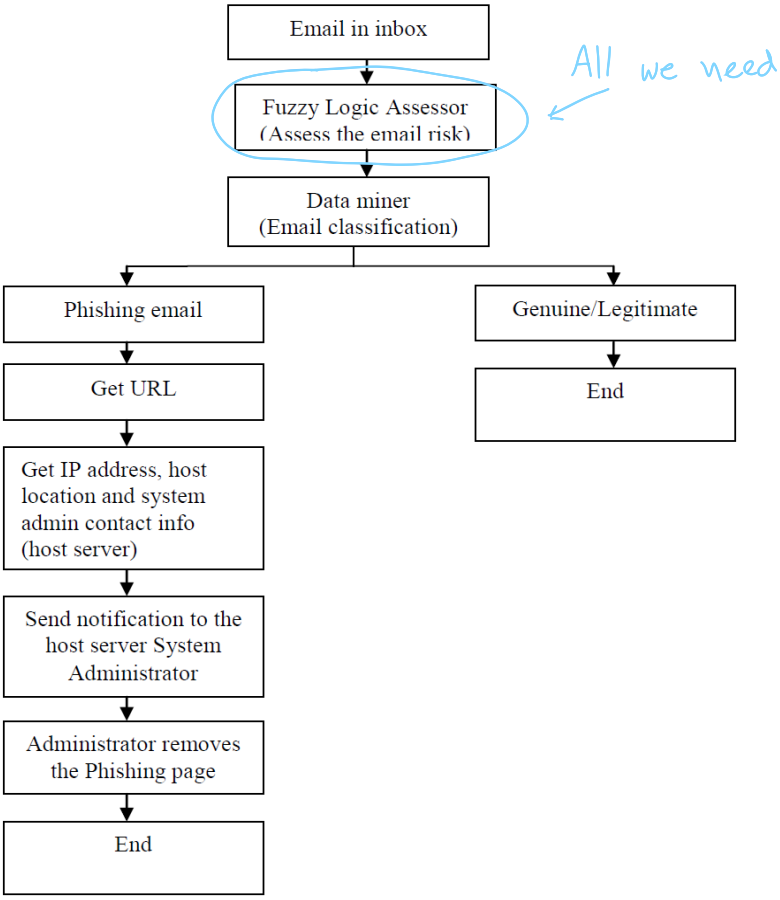


Figure 1. Overall system approach

3.2 Detecting and Classifying Phishing Email

A. Main Characteristics of Phishing Emails. The proposed methodology will apply fuzzy logic and data mining algorithms to classify phishing emails based on two classification approaches such as content-based approach and non-content based approach. Specific categories or criteria are selected for each approach. The components or selected features are then identified for each category. The list of the classification approaches with the identified criteria and specific features is listed in the table below. The list will be used as basis for in the simulation and determination of phishing emails.

The main characteristics of phishing emails are listed in Table 1.

Table I. Characteristics and stages of the components of phishing emails

Classification Approach	Category/Criteria	Component	Stage/Layer
Non-content Based Approach	URL	IP URL	Stage 1
		Redirect URL	
		Non-matching URL	
		Crawler URL	
		Long URl address	Weight = 0.5
		URL prefix/suffix	
Content-based Approach	Email Message	Spelling Errors	Stage 2
		Keywords	
		Embedded links	Weight = 0.5
Overall Weight			1.0

3.3 Use of Fuzzy Logic and RIPPER Data Mining Algorithm

The approach is to apply fuzzy logic and RIPPER data mining algorithm to assess phishing email based on the 9 identified characteristics or components. The essential advantage offered by fuzzy logic techniques is the use of linguistic variables to represent key phishing characteristic or indicators in relating phishing email probability.

A. Fuzzification. During fuzzification, 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 [7]. For example, redirect URL can range from „low" to „high" with other values in between. 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. Each point in the input space is mapped to a membership value between [0, 1]. For each input the values ranges from 0 to 6 while for output, the value ranges from 0 to 100. A plot of the membership function of the (Redirect URL) which is one of the phishing characteristic and the linguistics descriptors is illustrated in Figure 1 below. The same approach is used to calibrate the other 8 key phishing characteristic indicators.

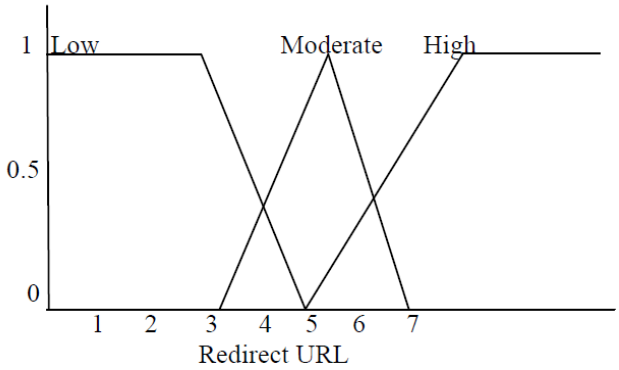


Figure 1. An example of the linguistic descriptors that will be used to represent one of the key phishing characteristic indicators and a plot of the fuzzy membership functions

B. Rule Generation using the RIPPER classification algorithm. This study proposes to use the RIPPER algorithm to learn the relationships of selected different phishing features. Such classification algorithm is selected to learn the relationships of the different phishing features. Fuzzy rules are represented as if-then statement which is based on the phishing email probability with the different phishing characteristics or indicators. Classification is done using WEKA.

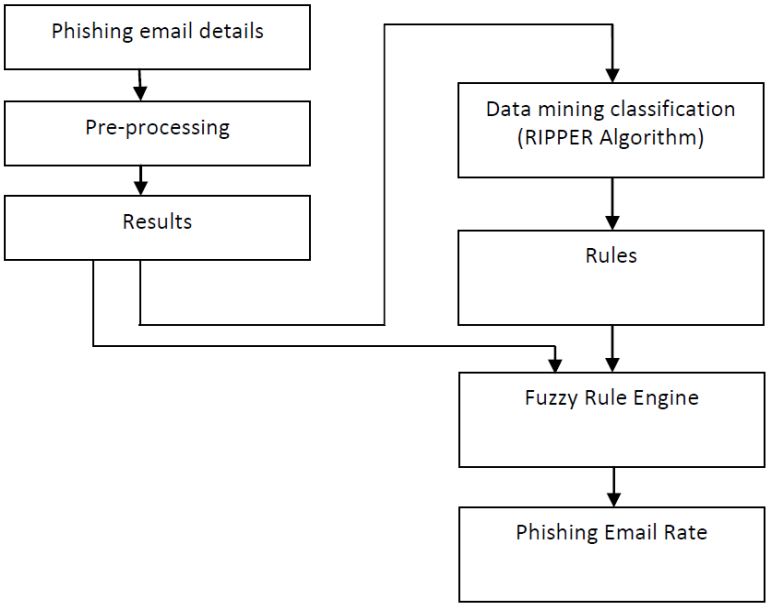


Figure 2. Determination of the Phishing email risk rate

C. Rule Outputs. The outputs of all discovered rules will be unified. Combining the membership function of all the rules ensuing consequent previously scaled into single fuzzy sets or outputs.

D. Deffuzification. Deffuzification is the process of producing a measurable result in fuzzy logic given the fuzzy sets and membership degrees. It is a process in fuzzy logic where valuable data is produced from fuzzy data. This process transforms a fuzzy output of a fuzzy inference system into a crisp output [12]. Fuzzification facilitates in evaluating the rules, but the final output has to be a crisp number. The input for the defuzzification process is the collective fuzzy set and the output is a number.

A useful defuzzification technique is the *center of gravity*. A typical set membership function has the graph of a triangle. The first step of defuzzification normally removes parts of the graph to form a trapezoid. The trapezoids are then superimposed one after the other to form a single geometric shape. The centroids which is called *fuzzy centroid*, is calculated. The *x* coordinate of the centroid is the defuzzified value. The output is the Phishing email risk rate and is defined in fuzzy sets as "Fraudulent" to "Genuine or Legitimate".

3.4 Fuzzy Data Mining Inference Rules

A. Rule Base for Stage 1 or Layer 1. There are six input parameters for the rule base and it has one output. It contains all the if-then rules of the system. In each rule base, every component is assumed to be one of three values (based on the linguistic descriptor) and each criterion has six components. Hence rule base 1 contains $3^6 = 72^9$ entries. The output of rule base 1 is one of the phishing email rate fuzzy sets (Genuine, Suspicious 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 sho wn in Table IV. The system structure for URL & Domain Identity criteria is the joining of its five components, which produces the URL & Domain Identity criteria (Layer one).

Non-content based

Table 2. Sample of the rule base stage 1 entries for the URL Domain and Entity Criteria

Rule #	IP URL	Redirect URL	Non-matching URL	Crawler URL	Long URL address	URL Prefix/suffix	URL Domain Entity & Criteria
1	Low	Low	Low	Low	Low	Low	Valid/Genuine
2	Low	Low	Low	Low	Low	Moderate	Valid/Genuine
3	Low	Low	Low	Moderate	Moderate	Moderate	Suspicious
4	Moderate	Low	Moderate	Low	Low	Moderate	Suspicious
5	Low	Low	Low	Moderate	Moderate	High	Suspicious
6	Moderate	Moderate	Moderate	High	High	High	Fraud
7	High	High	High	High	Moderate	Moderate	Fraud

B. Rule Base for Stage 2

Table 3. Sample of the rule base stage 2 entries for Email Content Domain

Rule #	Spelling Errors	Keywords	Embedded Links	Email Content Domain
1	Low	Low	Moderate	Genuine
2	Low	Moderate	Moderate	Suspicious
3	High	High	High	Fraud
4	Low	Low	Low	Genuine
5	High	Moderate	Moderate	Fraud
6	Moderate	Low	Moderate	Suspicious
7	Moderate	Moderate	Low	Suspicious

Content based

3.5 Locating the Host Server of the Phishing Page

WHOIS is a protocol used to find information about networks, domains and hosts. The WHOIS query is used to locate the host server of a phishing page. WHOIS is a query/response protocol that is widely used for querying an official database. The WHOIS database contains IP addresses, autonomous system numbers, organizations or customers that are associated with these resources, and related Points of Contact on the Internet [4]. A WHOIS search will provide information regarding a domain name, such as example.com. It may include information, such as domain ownership, where and when registered, expiration date, and the name servers assigned to the domain.

A WHOIS server listens on Transmission Control Protocol port 43 for requests of the host server and related contact information sent through web-based referrals [4]. Its connection is closed as soon as the output is finished. The closed connection is the indication to the client that the response has been received. The system runs the WHOIS query on the URL that is contained in the Phishing email.

3.6 Removal of the Phishing Page

Upon receiving the notification of the phishing page"s existence on the host server, the hosting administrator will then test the legitimacy of the phishing link and its validity. Once the Administrator confirms the phishing page, the infected or hacked website will be shut down immediately to protect Internet users from further phishing. The host Administrator then notifies the website owner about the existence of the phishing page within their website. As soon as the phishing page is removed, if no notification has been sent, the proposed system will periodically check for evidence that it has been removed.

This technique assumes that website owner and host Administrator are absolutely unaware of the presence of the phishing page within their website or server until our

technique notifies them. This means Phishers are taking control of the legitimate website to upload their phishing page. By doing so phishing pages are able to bypass anti- virus securities installed on the user's computer.

4 Results

Publicly available datasets from Phistank were used for simulation. There are two stages in determining the fuzzy data mining inference rules. 1000 sample instances are used from the Phistank archive. For rule base 1, there are 6 identified Phishing email characteristics based on the non-content based approach. The assigned weight is 0.5. For rule base 2, there are 3 identified characteristics of Phishing emails based on the content-based approach. The assigned weight is 0.5. The email rating is computed as 0.5 * URL and Domain Entity crisp (rule base 1) + 0.5 * Email Content Domain crisp (rule base 2).

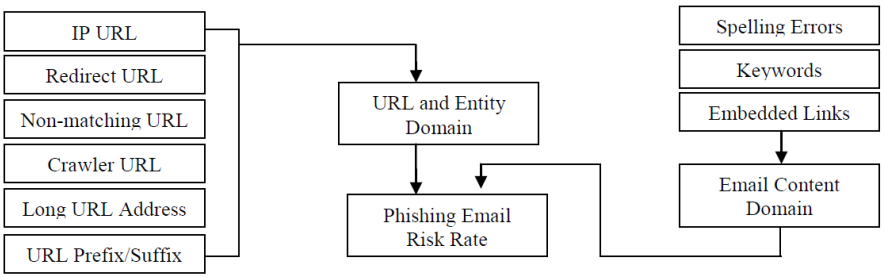


Figure 3. Fuzzy logic inference overall system use in assessing the Phishing email risk rate

The study utilizes the RIPPER data mining algorithm. The RIPPER algorithm uses separate and conquer approach. It is considered an inductive rule learner that builds a set of rules that identify the classes while minimizing the amount of error. The error is determined by the number of training examples that are misclassified by the rules. The prediction accuracy is recorded in Table 4.

Table 4. Results generated from the WEKA classifier using RIPPER algorithm applied to classify Phishing emails

Validation Mode	10 fold cross validation
Attributes	URL Domain and Entity Criteria
	Email Content Domain
Number of rules	12
Correctly classified	85.4%
Incorrectly classified	14.6%
Number of samples/instances	1000

True Pos

False Pos

The initial results showed that URL and Entity Domain and the Email Content Domain are important criteria for identify and detecting Phishing emails. If one of them is "Valid or Genuine", it will likely follow that the email is a legitimate email. The same is true if both of the criteria are "Valid or Genuine". Likewise, if the criteria are "Fraud", the email is considered as a Phishing email.

The results regarding how many Phishing sites were removed as a result of the notifications sent after the email is classified as a Phishing email is yet to be established. Moreover, the final results for the input combinations shown in Table 1 and Table 2 is still to be determined. At this stage, data gathering is still on-going.

5 Conclusion

URL and Entity Domain as well as Email Content Domain are two important and significant Phishing criteria. If one of the criteria is "Valid or Genuine", it will likely follow that the email is a legitimate email. The same is true if both of the criteria are "Valid or Genuine". Likewise, if the criteria are "Fraud", the email is considered as a Phishing email. It should be noted, however, that even if some of the Phishing email characteristics or stage is present, it does not automatically mean that the email is a Phishing email. The initial objective is to assess the risk of the email in the archive data using fuzzy logic and the RIPPER classification algorithm. Several characteristics were identified and major rules that were determined along the study were used in the fuzzy rule engine.

Moreover, the number of Phishing sites removed as a result of sending a notification to the host server as soon as the email is classified as Phishy should be monitored to determine the effectiveness of the study. There may be instances where the host server will not remove the Phishing site even if it has received notifications. This case is beyond the context of the study as this requires administrative sanctions or inclusion of the site in the Blacklist. The study however was able to prove that fuzzy logic and data mining with the use of the RIPPER algorithm is in a way sufficient in assessing the risk of a Phishing email and classifying the email as such, thereby resulting in the issuance of notification to the host server for removal of the Phishing page.

References

[1] A.-P. W. Group, "Global phishing survey: Domain name use and trends in 2h2010,"http://www.antiphishing.org/reports/APWG, GlobalPhishingSurvey 2H2010.pdf.
[2] A. P. W. Group, "Phishing activity trends report," 2009, http://www.antiphishing.org/reports/apwg reportQ42009.pdf.
[3] Anti Phishing Working Group, "Phishing Activity Trend Report", Jan-March 2008

[4] Shah, R.; Trevathan, J.; Read, W.; Ghodosi, H.,"A Proactive Approach to Preventing Phishing Attacks Using the Pshark Model", IEEE Sixth International Conference on Information Technology: New Generations, March 2009, pp. 915 - 921
[5] Afroz, S.; Greenstadt, R., "PhishZoo: Detecting Phishing Websites by Looking at Them", IEEE Fifth International Conference on Semantic Computing (ICSC), 2011, pp. 368-375
[6] Sophos, "Security Threat Report", July – 2008
[7] Maher Aburrous, M. A. Hossain, Keshav Dahal, Fadi Thabatah, "Modelling Intelligent Phishing Detection System for e-Banking using Fuzzy Data Mining", IEEE International Conference on CyberWorlds, 2009, pp265 - 272
[8] 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.

[9] Rokach, Lior; Oded Maimon (2008). *Data mining with decision trees: theory and applications*. World
Scientific Publishing. ISBN 978-9812771711.

[10] L. James, "Phishing Exposed," Tech Target Article sponsored by: Sunbelt software, searchexchange.com, 2006 [11] http://www.phishtank.com/phish_archive.php
[12] J. Milletaty, "Technical Trends in Phishing Attacks", *US-CERT, http://www.uscert.gov/reading_room/*

[13] T. Sharif, "Phishing Filter in IE7,"http://blogs.msdn.com/ie/archive/2005/09/09/463204.aspx September 9, 2006.

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

[15] J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Identifying Suspicious URLS: An Application of Large-scale
Online learning," in *ICML '09:Proceedings of the International Conference on Machine Learning*,2009, pp.
681–688.