DEMONSTRATING DIFFERENT PHISHING ATTACKS USING FUZZY LOGIC

Ms.Shweta Dasharath Shirsat
IT Department
M.L. Dahanukar College of Commerce
City: Mumbai Country: India
Email: shwetadshirsat@gmail.com

Abstract-**Phishing has increased tremendously over last few years and it has become a serious threat to global security and economy. Existing literature dealing with the problem of phishing is scarce. Phishing is a deception technique that uses a combination of technology and social engineering to acquire sensitive information such as online banking passwords, credit card or bank account details[2]. Phishing can be done through emails and websites to collect confidential information. Phishers design fraudulent websites which look similar to the legitimate websites and lure the user to visit the malicious website. Therefore, the users must be aware of malicious websites to protect their sensitive data[1].But it is very difficult to distinguish between legitimate and fake website especially for nontechnical users [4]. Moreover, phishing sites are growing rapidly. The aim of this paper is to demonstrate phishing detection using fuzzy logic and interpreting results using different defuzzification methods.**

Keywords-Fraudulent,Defuzzification,Membership functions,PageRank,Indexing

I. INTRODUCTION

Phishing is the fraudulent practice of acquiring confidential data through the fraud website which appears to be legitimate. Detection of phishing websites is difficult because it is very easy for an attacker to create an exact replica of the original website which looks convincing to the user. Typically, a phishing attack starts by sending an Email to an authenticated user and convincing him to visit a URL link given in the E-mail[4]. The content of the URL link may ask the user to enter some bank details or any confidential information which might be useful for an attacker to perform further attacks. Thus to protect an individual from such malicious practices, there should be the robust mechanism. Such a mechanism is represented in this paper which uses fuzzy logic to determine the state of a website using predefined parameters.

II. LITERATURE REVIEW

The following section describes the literature review as:

(K.N.Manoj Kumar,et.al.2015) [1] in their paper they explained how to classify and predict phishing websites. The research methodology used is a fuzzy rule-based model.This framework works better and gives a lower error rate.

(Luong Anh Tuan Nguyen, et.al.2015)[2] discussed improving detection of phishing websites using Neuro-fuzzy network model. Their experimental results show 99.22% accuracy which is 2.95% higher than the existing technique.Their future plan is to enhance the proposed system using larger datasets and more heuristic parameters.

(Rajeev Kumar Shah, et.al. 2016)[3]discussed developing an improved phishing email classifier with better prediction accuracy and fewer numbers of features. Their methodology used is Fuzzy Logic (FL) model with association classification mining algorithms. It is noted that the proposed tool offered the best performance among the tested tools, being about 11% better compared to Netcraft and 6% better compared to Spoofstick.

(P.A.Barraclough,et.al.2014)[4}this study presents a parameter tuning framework, using adaptive Neuro-fuzzy inference system with comprehensive data to maximize systems performance. They conclude that this framework works better and gives a lower error rate and an accuracy of 98.74%. The work does be done next is to extract large data from a wide range of samples and use different cross-validation with large data-sets.

(Ms.Roshni Vitthal Pawar, et.al.2017)[5]this paper proposed a neuro-fuzzy model without using rule sets for phishing detection tools. Their paper concluded by stating that 95.00% accuracy is provided using the proposed technique.

(Luong Anh Tuan Nguyen, et.al 2016)[6]This paper presents a novel approach to overcome the difficulty and complexity in identifying phishing sites. The results show that the proposed technique can identify with an accuracy identification rate of above 99%.In the future, this neuro-fuzzy model can be improved to enhance the identification ratio.

(Anugrah Kumar, et.al.2016)[7]this paper proposes an approach towards Phishing Detection Using Rough Set Theory. The Limitations of this approach is that it only determines the probability of a site to be reliable or unreliable.

(Sadia Afroz, et.al.)[8]their paper proposes a phishing detection approach—Phish Zoo—that uses profiles of trusted websites for detecting phishing attacks. Their future scope is to create a robust system for phishing detection with minimal human intervention.

(Anindita Khade, et.al.2013) [9]Detection of Phishing Websites Using Data Mining Techniques" this paper we propose a method which combines fuzzy logic along with data mining algorithms for detecting phished websites. The results showed that the RIPPER algorithm achieved 85.4% for correctly classified Phishing emails and 14.6% for wrongly classified Phishing emails. The phishing page removal success rate is 81.81%.

(Phoebe Barraclough, et.al.2015)[10]This study investigates and identifies parameters in a single platform based on fuzzy system and neural network for phishing websites detection. the result suggests that fuzzy systems and proper parameter tuning together with wide-ranging effective data can detect phishing websites with a higher accuracy. Their future work will be to extract a large data and utilize 20-fold cross-validation to measure the model's accuracy.

(Mohammed Nazim Feroz et.al.2015)[11] This paper proposes a system capable of clustering, classifying, categorizing, and ranking URLs in real-time while adapting to new and evolving trends in URL characteristics.The methodology used is URL page ranking algorithm. Their classification accuracy of the is 98.46%.

(Rosana J. Ferolin et.al.)[12]their objective is to assess the risk of the email in the archive data using fuzzy logic and the RIPPER classification algorithm. The research methodology used Fuzzy logic and RIPPER data mining algorithm. The study was able to prove that RIPPER algorithm provides accurate and best results as compared to other data mining algorithms.

(Ms.S.Nivedha et.al.2017)[13]the proposed application concentrates on the fuzzy value classification based on the apriori rule applied. After the association rule applied, the performance evolution is done for the association rule applied and binary matrix generated values. The upcoming developments that could be expected is Fuzzy based association rule mining

(Mona Ghotaish Alkhozae, et.al.2017)[14]this paper, they proposed a phishing detection approach based on checking the webpage source code.They have used a phishing website detection methodology. Their future work is to add other checks in the program and check more source codes contains many languages in it like PHP, CSS, asp, Java, Perl, etc.

(Phoebe Barraclough, et.al.2017)[15]this paper contributes by constructing a fuzzy rule model using a combined effective feature-set that has shown an excellent performance. The research methodology used is rule-based feature-driven cyber phishing detection system based on ANFIS. This model gives an accuracy in the range of 94.66% - 99.2%.

III. PROPOSED STUDY

Fuzzy logic is a is a technique for handling imprecise and vague information. It is a rule-based system which consists of a set of if-then rules. In the fuzzy system, values are indicated by a number from range 0 to 1 where 0 represents absolute falseness and 1 represents absolute truthfulness[16].In this paper, we have demonstrated the use of fuzzy logic to determine whether the given webpage is phished or legitimate. It classifies the web pages depending on the set of predefined rules.There are different phishing website detection rules through which a user can decide whether the website is legitimate or suspicious. Some of them are listed below[13]:

1. Address Bar based Features

a. Length of URL

Rule:IF
{URL length<54--->feature=Legitimate else if URL length>=54 and<=75--->feature=Suspicious
Otherwise ---->feature=phished}

b.Using URL Shortening Services

Rule:IF
{TinyURL--->phished Otherwise---->Legitimate}

c. URL's having "@" Symbol

Rule: IF
{URL having @ symbol---->Phished Otherwise ---->Legitimate}

2. Domain based Features

a.Domain Age

Rule: IF
{Age of domain >=6 months---->Legitimate Otherwise ---->phishing

b. DNS Record

Rule:IF
{no DNS record for the domain---->Phishing Otherwise----> Legitimate}

c. Website Traffic

Rule:IF
{Website Rank<=100,000---->Legitimate Website Rank>100,000---->Suspicious Otherwise ----

>Phished}

d. PageRank

Rule:IF

{Pagerank<0.2---->Phishing Otherwise ---->Legitimate}

e. Google Index

Rule:IF

{Webpage indexed by google---->Legitimate Otherwise ---->Phishing}

3. HTML and JavaScript based Features

a.Status Bar Customization

Rule:IF

{onMouseOver changes Status Bar---->Phishing It doesn't change Status Bar---->Legitimate}

b.Disabling Right Click

Rule:IF

{Right click disabled---->Phishing Otherwise ---->Legitimate}

c.IFrame Redirection

Rule: IF

{Using iframe---->Phishing Otherwise ---->Legitimate}

The above set of rules is divided into 3 main categories: address bar based features, domain based features, HTML and javascript based features. Where the domain based features identify the authenticity of the website, HTML, and javascript based features maintain the integrity of the website and address bar based features provides reliability to the webpage. This paper presents an approach to quickly detect phishing websites using Fuzzy Logic. The approach is based on some characteristics that are present in the website.Depending on these rules we can conclude whether the given webpage is Highly Legitimate, Legitimate, Suspicious, Phishy, Highly Phishy.

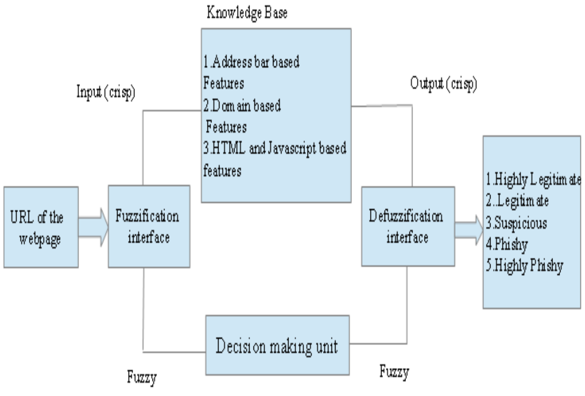

Fig 1.URL Phishing detection system

The above model consists of following steps:

Step 1: Initially, in the fuzzification unit, the crisp input is converted into the fuzzy input.

Step 2: Determine the set of fuzzy rules. A rule base contains numerous fuzzy IF-THEN rules

Step 3: A database defines the membership function of fuzzy sets used in fuzzy rules. The database and rule base are collectively called knowledge base. For this demonstration, the rules are divided into 3 categories: Address bar based features, Domain-based features, HTML and Javascript based features.

Step 4: Finally, defuzzification process is carried out to produce crisp output. Then, suitable decisions are made in the decision making unit.

IV. RESULT AND ANALYSIS

The training dataset contains 300 random set of URLs from phishtank and DMOZ. The demonstration is based on a fuzzy logic approach using triangular membership function and then used three different defuzzification methods as given below:

1. Mean of maximum

In MOM defuzzification method, the fuzzy logic controller first identifies the scaled membership function with the greatest membership. the output is given by[17],

$$x^* = \frac{\sum_{i=1}^{n} \bar{x}_i}{n}$$

Using this method it is found that in the given dataset contains: 50 Highly Legitimate URLs, 48 Legitimate URLs, 60 Suspicious URLs, 92 Phished URLs and 50 Highly Phished URLs.

2. Weighted Average method

This method is valid for symmetrical output membership function only. Each membership function is weighted by its value. The output is given by[17],

$$x^* = \frac{\sum \mu_C(\bar{x}_i) \cdot \bar{x}_i}{\sum \mu_C(\bar{x}_i)}$$

Using this method it is found that in the given dataset contains: 60 Highly Legitimate URLs, 56 Legitimate URLs,56 Suspicious URLs,78 Phished URLs and 50 Highly Phished URLs.

3. Centroid method

In this method, the results of the rules are added together to generate specified output. It also is known as the center of mass, the center of the area, center of gravity method. The output is given by[17],

$$x^{*} = \frac{\int \mu_{C}(x) \cdot x dx}{\int \mu_{C}(x) dx}$$

Using this method it is found that in the given dataset contains: 52 Highly Legitimate URLs, 74 Legitimate URLs,56 Suspicious URLs,78 Phished URLs and 40 Highly Phished URLs.

The results of phishing detection system are as follows[17],

URLs	**Defuzzification methods**		
	Mean of maximum principle	**Weighted average method**	**Centroid method**
http://facelook.shop.co/login.php	Phished	Highly phished	Highly phished
http://www.esmartstart.com	Highly phished	Phished	Phished
http://facebooook.axfree.com/	Suspicious	Legitimate	Suspicious
https://paytm.com/	Highly Legitimate	Legitimate	Highly Legitimate
https://www.amazon.in/	Legitimate	Highly Legitimate	Legitimate

Table 1.Results of Defuzzification methods

V. CONCLUSION

Detecting malicious URLs is one of the major problems on the internet[2].This paper demonstrates the detection of phishing website using the fuzzy logic approach with five heuristic parameters(Highly Legitimate, Legitimate, Suspicious, Phished and Highly Phished).This technique is successfully implemented and tested on a dataset containing 300 different URLs. Our future scope is to enhance the system by using larger datasets and more heuristic parameters.

ACKNOWLEDGEMENT

I would like to express my special thanks of gratitude to Prof.Dr.D.R.Kalbande Dean (Industry Relations)Professor & Head Department of Computer Engineering Bharatiya Vidya Bhavan's Sardar Patel Institute of Technology for providing their invaluable guidance,comments and suggestions throughout the research paper work.

REFERENCES

[1]K.N.ManojKumar,K.Alekhya,"Detecting Phishing Websites using Fuzzy Logic" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 10, October 2016.

[2] Luong Anh Tuan Nguyen and Ba Lam To and Huu Khuong Nguyen and Chauan Pham and Choong Seon Hong "A Novel Neuro-Fuzzy Approach for Phishing Identification" , Journal of Automation and Control Engineering Vol. 3, No. 6, December 2015.

[3]Rajeev Kumar Shah,Md.Altab Hossin,Asif Khan "Intelligent Phishing Possibility Detector" International Journal of Computer Applications (0975 – 8887)Volume 148 – No.7, August 2016.

[4]P.A.Barraclough,G.Sexton,M.A.Hossain,N.Aslam "Probustarameter optimization for intelligent phishing detection using Adaptive Neuro-Fuzzy" (IJARAI) International Journal of Advanced Research in Artificial Intelligence,Vol. 3, No.10, 2014 Extended Paper from Science and Information Conference 2014.

[5]Ms.Roshni Vitthal Pawar,Ms. Ruchita Pandit Rao Pawar,Ms.Pranali Ganesh Salunkhe and Ms.Ankita Gajanan Sankhe, "Phishing Identification Using An Efficient Neuro-Fuzzy Model", Volume 2,Issue 4, April 2017,ISSN (online): 2456-0006 International Journal of Science Technology Management and Research.

[6]Luong Anh Tuan Nguyen,Huu Khuong Nguyen ,"Phishing Identification Using a Novel Non-Rule
Neuro-Fuzzy Model" (IJCSIS) International Journal of Computer Science and Information Security,
Vol. 14, No. 4, April 2016.

[7] Anugrah Kumar, Sanjiban Shekar Roy, Sankalan Saxena, Sarvesh SS Rawat, "Phishing Detection by determining Reliability Factor using Rough Set Theory".

[8] Sadia Afroz,Rachel Greenstadt "Phish Zoo: Detecting Phishing Websites By Looking at Them"

[9]Anindita Khade,Dr.Subhash.K.Shinde,"Detection of Phishing Websites Using Data Mining Techniques",International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 12, December - 2013 IJERT ISSN: 2278-0181.

[10] Phoebe Barraclough, Graham Sexton "Phishing Website Detection Fuzzy System Modelling"Science and Information Conference 2015 July 28-30, 2015 | London, UK.

[11]Nazim Feroz, Susan Mengel "Phishing URL detection using URL Ranking"2015 IEEE International Congress on Big Data

[12]Rosana J. Ferolin "A Proactive Anti-Phishing Tool Using Fuzzy Logic and RIPPER Data Mining Classification Algorithm "

[13]Ms.S.Nivedha,Mr.S.Gokulan,Mr.C.Karthik,Mr.R.Gopinath,Mr.R.Gowshik"Improving Phishing URL Detection Using Fuzzy Association Mining " The International Journal of Engineering and Science (IJES) Volume 6, year 2017 || ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805.

[14]Mona Ghotaish Alkhozae, Omar Abdullah Batarfi "Phishing Websites Detection based on Phishing Characteristics in the Webpage Source Code"Volume 1 No. 6, October 2011 ISSN-2223-4985 International Journal of Information and Communication Technology Research.

[15]Phoebe Barraclough, Gerhard Fehringer “Intelligent Detection for Cyber Phishing Attacks using Fuzzy Rule-Based Systems” International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization)Vol. 5, Issue 6, June 2017 .

[16]Rami M. Mohammad,Fadi Thabtah,Lee McCluskey “Phishing Website Features” School of Computing and Engineering, University of Huddersfield, Huddersfield, UK.

[17]S.N.Sivanandam,S.N.Deepa “Principles of Soft computing” Second edition, Wiley India.