ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 10, October 2016

Detecting Phishing Websites using Fuzzy Logic

K. N. Manoj Kumar, K. Alekhya

***Abstract*— Phishing websites contains malicious web pages designed by fraudulent people and attempt to loot personal information from users. Phishing emails are the messages designed by the particular phishing website sent to the users in order to gain passwords, account and credit card details etc. These phishing websites and emails harm their victim by plundering the identity and money. Due to this people may lost their trust in Internet transactions. Detecting and removing the phishing websites is really a complex problem because every time the phishes are approaching with new techniques and methodology. This phishing is done through websites as well as emails. Many anti-phishing techniques have been developed and many researchers proposed their algorithms to thwart the phishing attacks. This paper discussed various phishing techniques and algorithms developed by the researches and outlined their merits and demerits. This paper also discuss the effective approach in dealing with phishing sites by gaining the advantage from the genetic algorithm to deal with fraudulent websites and implement through fuzzy logic technique.**

***Index Terms*— fuzzy logic, phishing websites criteria, phishing, risk assessment.**

I. INTRODUCTION

. Phishing is an online criminal act of looting personal information, bank and credit card details through internet by sending fake emails from fraudulent websites. This is a web based attack done by phishers. [2, 3]They carry out the fraudulent transactions and acquire confidential messages on behalf of the user (victim) with the help of information stolen from them. [4]The impact of phishing is very drastic leads to identity theft and financial losses. [8] As they are gaining advantage on the increased use of online day to day services such as online banking and shopping.

Now a days there is an increase in the fraudulent websites. These websites looks exactly as the original websites so that the users easily get attacked [1]. In early days phishing attacks is done through e-mails. The spoofed e-mails is sent to many users by the attackers so that the victims can sent back sensitive information like usernames and passwords. [7] Present days these type of attacks has very low success rate because the user has learnt not to send such confidential information.

K.N.Manoj Kumar, B-Tech Information technology site school, VIT Universirty, Vellore, India, Mobile No. : 9962419392.
K. Alekhya, B-Tech Electronics and Communication Engineering, AN University, Guntur, India, Mobile No.: 8978518829.

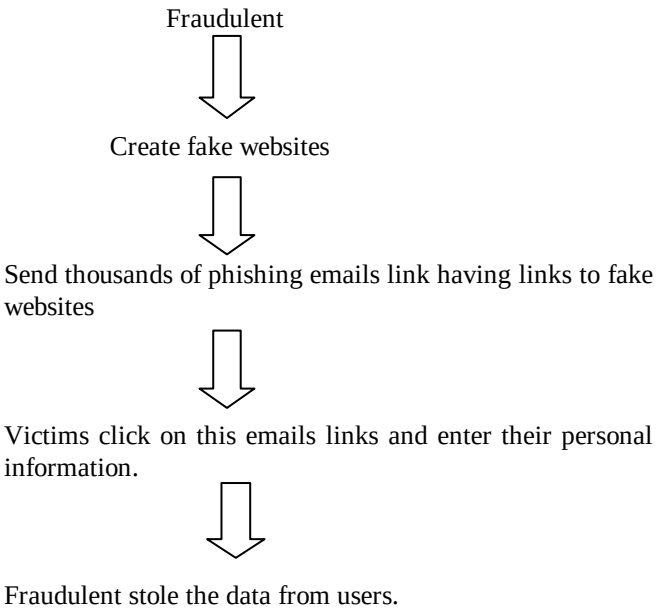

Fig 1. Process of phishing a website

Many bank organizations does not provide interactive services via e-mails because the users has to give passwords. These organizations provide services through websites so that they can depend on encryption technology using SSL protocol. These days phishers are gaining information by taking the advantages of the vulnerabilities in the browsers like Internet Explorer etc. [7] These browsers pop up some window asking about the bank or other personal information so that they can easily gain information from users. The success rate is very high in this type of fraudulent websites because of the reason they appears as the original websites.

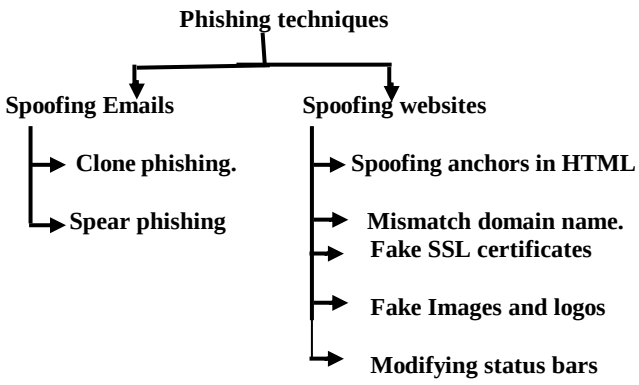

Fig 2. Shows different types of phishing techniques.

II. RELATED WORKS

These are some of the ways where phishing is generated. [3] To minimize the phishing attacks many solutions came into existence. Some of them are generic algorithm, visual similarities based anti phishing and String matching algorithm. Generic algorithm has a set of rules which detects whether a website is fraudulent or not it.[2] It acts like if else condition if the URL of the website matches with the rules set in the generic algorithm then it is an original website else it is phished one. Visual similarity based model compare the logos, signature of original website with the fraudulent website so that they calculate the distance between both sites and hence detect phishing site. String matching algorithm divide the URL into several tokens and check the amount of similarity between two URL and if there is and difference in tokens it is decided as a fake URL. Many efforts have been made to compare the performance of some machine learning techniques like fuzzy logic and neural network theory to easily detect spam emails. However,[7] these attempts still need improvement to get a higher success rate.

As we consider the growing in technology phishers are coming with new techniques so that the existing techniques may not work properly [2]. In logo detection technique the phisher may attach exactly same image as the original website [2]. There are some counter measures to defeat phishing. Kirda et al [11]. created an extension to browser which protect users from the spoofed websites. Spoofed Guard, Net craft, Phish Zoo and Site Advisor are some of the toolbars developed to warn users about phishing attacks[11]. Phishing is very challenging task when comes under multi-national companies. Some MNC companies like Google, Microsoft, PayPal uses different approaches to keep their uses to browse safely without any phishing attacks. Browser like chrome which are maintained by Google alerts the users if they are trying to visit webpages which are suspected of phishing or malware. When the phishing and malware detection alerts in a browser is turned on you might see the following messages like The website a head contains malware!, Deceptive site ahead, the site ahead contains harmful programs.

Various works have been done before to stop phishing attacks on websites and links. In this section we will see a detailed review over the previous work.

Recently researchers [7] had implemented a technique which is based on neuro fuzzy method. This method uses five inputs (popup from mails, phish tank, user behavior outline, user specified sites, justified site rules) to classify phishing site with more precision and accuracy based on two fold cross validation. Several research techniques are single folded protection which leads inaccuracy in results. A sum of 250 with these five inputs used in testing giving very promising results as compared to other previous results in this field.

Another genetic algorithm [8] based technique which is employed to differentiate between legitimate and phishing links. This evaluation is done by crossover, estimation function and mutation. This works on the basis of the rule set which is stored in database that counter parts the phishing links by matching the URL with this rule set. If the link counterparts every rule in the database then it is operating by fraudulent. This approach is sufficient to perceive the phishing links with very minimal negativities.

This is the navel approach [9] which overcomes the intricacy and complexity in detecting the phishing websites. This is the beneficent method based on categorization and association data mining algorithms which optimized with PSO(particle swarm optimization) algorithm. using this algorithm we can able to distinguish and classify all rules and factors to detect phishing websites. To categorize the authenticity of phishing training data sets this approach utilize MCAR classification algorithm and later optimized with ant colony optimization algorithm. but this approach has limitations like random decision making and time to convergence are uncertain in classification. To nullify these limitations we choose PSO algorithm to optimize this problem and this also improves the correct classification of phishing websites. this anti phishing detection project use JAVA technology.

Other method [10] which is knowledge base compound scheme based on parsing methods and inquiry operations that counters the phishing and other internet attacks by means of browsers. This technique examine the URLs before visiting the site which offers security from web based attacks. Also uses query processing and different parsing operation to distinguish various web attacks as well as phishing attacks. This technique merely effect the browser speed and absolutely based on browser. Using this methodology a browser can easily identify the phishing attacks, attacks based on hacking and SSL attacks. This method which deployed in browser can achieve 97% accuracy on fraudulent attacks.

The proposed method [11] PHISHZOO is a phishing detection method uses the trusted websites profiles to detect attacks. This offers similar accuracy to other methods like blacklisted approach. The advantage in using this approach is that it can categorize various phishing approaches and attacks on smaller websites(Intranets). This method also comprises performance analysis and framework for computer vision techniques.

This proposed work [12] which uses trusted mechanism for mutual authentication repeals towards man in the middle attacks, trust on wonderful user behavior and also guard users account even in the presence of spyware and key loggers. They also implemented the prototype and demonstrated practicality of system.

III. PROPOSED MODEL

Fuzzy logic was first introduced by Zadeh in 1965 as modification of classical set theory. Fuzzy set theory process the imprecise information by means[10] of membership function. Fuzzy logic allow the intermediate degree between notations such as true and false, hot and cold, black and white etc. as used in Boolean logic. In fuzzy system, values are indicated by a numbers from range 0 to 1. where 0 represents absolute falseness and 1 represents absolute truthfulness. Fuzzy logic is used to evaluate the degree of phishing in a variety of web pages we come across. It classifies pages based on the degree of phishiness present in the pages .Hence when

we employ fuzzy logic in order to detect phishing present in various webpages [11] it classifies webpages based on a certain set of metrics and helps us determine the degree of phishing. This is determined by employing a set of pre-defined rules.

If the URL of the given link matches with the rules given in the rule base then given input link violates the genuine website policies and hence considered as malicious and given a score. Here accordingly we consider 10 sets of rules which are used to evaluate the phishiness in the URL. If the 10 rule sets matches with the given input URL then it is considered as highly phishy. In the same way for the given input if any of those rules does not matches with the URL then it considered as a highly legitimate. The score for the given site is assigned from 0 to 1 based on the intensity of phishiness in the URL. Advantage in fuzzy logic is it determine the phishiness of the URL from less suspicion to high suspicion. Website phishing risk rate is assigned as highly Legitimate, Legitimate, suspicious, phishy and highly phishy and phishing character indicators as low, medium and high. When we compared with other approaches fuzzy logic approach is most efficient in detecting the phishing site as well as the intensity of phishiness in these sites. This approach categorizes the given input(URL) of the website into five categories like highly Legitimate, Legitimate, suspicious, phishy and highly phishy. Based on this categorization we can easily the define the website reliability. It uses very less memory when compared with other techniques and its inference speed is high. But the results obtained from this model is not 100% accurate and designing this model is slightly complex.

Why using fuzzy logic?

Fuzzy logic has been using from many decades in engineering and researches to embed the inputs into computer model for many applications. Fuzzy logic is mainly useful for people who involves in research and development. Fuzzy logic provide information which provide information to access, manage, and categorize the website phishing risks than the previous approaches. The importance of fuzzy logic is the use of the linguistic variables to represent the phishing indicators.

In this paper our approach is to detect the website phishing through fuzzy logic techniques. To detect the phishing websites using fuzzy logic technique there are four phases 1)Fuzzification 2)Evaluating rule 3)Aggregating the rule outputs 4)Defuzzification. Using some phishing characteristic indicator and website phishing probability. We can determine whether the URL from the website is legitimate or not.

1) Fuzzification

This is the first step in fuzzy logic process which involves in domain transformation where crisp inputs are converted into fuzzy inputs. This approach assess website phishing risk rate on the basis of 20 characteristics to stamp as forged website. The main advantages of fuzzy logic is use of linguistic variables to represent key phishing characteristic and website phishing probability rate and these linguistic descriptors such as High, Low, Medium are assigned to range of values for each key phishing character. Valid website URL's are taken as an input and divided into classes. These classes are also called as fuzzy sets. We can't specify the clear boundary between the classes. The degree of presence of the values of the variables in a selected class is class the degree of membership. For each phishing character indicator membership function which is a curve how each point in input is mapped to membership value between the value 0 and 1. Website phishing rate risk is assigned as highly Legitimate, Legitimate, suspicious, phishy and highly phishy and phishing character indicators as low, medium and high. For example a long URL address is taken to represent the phishing character indicator and plot the membership function. By representing with fuzzy logic is closely related to human cognition

2) Evaluating rule

After specifying the phishing character indicators the next step is to find website phishing probability varies as a function with phishing character indicators. So we used **if….then** statements to evaluate inputs with the phishing character indicators so that the website phishing probability varies. The statement if….then works as if the input satisfy with the given phishing character indicator it is treated as legitimate website otherwise it is a fraudulent website. The condition acts as

If(condition)

Then

(act)

Example:-If(if the ip address of the received e mails matched with the rule)

Then

(phishing email)

We can characterize website phishing by some factors like web surveys, anti-phishing tools analysis, website phishing experiments.

3) Aggregating the rule outputs.

This is process in we unify or adding the outputs of all the rules. Combining the membership functions of all the outputs of the rule into single output(single fuzzy set).

4) Defuzzification.

It is the final process in fuzzy logic. It is a process in which we can transform the fuzzy output into crisp output. With the help of fuzzy logic we try to evaluate the rules but the final output must be a crisp value which mean that the website is either a phishy or legitimate one.

Website phishing risk rate is a fuzzy output and it varies from 'very phishy' to 'very legitimate'. Then the fuzzy output is defuzzified into scalar value.

IV. ALGORITHM FOR FUZZY EVALUATION

1. In the first step take an input (URL) continue with the first phase until linguistic variables and key phishing characters are set.
2. Loop until no key phishing character is left to evaluate the input.
    a) Evaluation is done by **if…then** rule and the output for each key phishing character is noted.
    b) Unify the outputs of all the rules into single output or single fuzzy set.
    c) Evaluate the website phishing risk from the above fuzzy output.
3. Transform the fuzzy output into crisp output [0,1] either 1 or 0.
4. From the value of crisp output determine the input as phishy or legitimate one.
5. Continue step 1, 2, 3, 4 for every input to be evaluated then save and quit the process.

layer No:	RULE BASE	Layer weightage
1	Using lengthy URL as a link. Using IP address rather than DNS name.	**0.1**
2	Large Number of Dots in IP Address. Using Modified Port Number.	**0.1**
3	Suspicious SSL Certificate. Age of Domain is Less than 6 months.	**0.1**
4	Unsecured Page and Redirected Pages. Taking Longer time to access accounts.	**0.1**
5	Taking Longer time to access accounts. Using Java Scripts to hide information.	**0.1**
6	Visual similarity to other pages. URL present in Google's Blacklist.	**0.1**
7	Using forms with submit button. Using pop-up windows.	**0.1**
8	Much importance on security and response. Using more time to access accounts.	**0.1**
9	Server form handler(SFH). Using mouse over to obscure the link.	**0.1**
10	Adding prefixes and suffixes in web address bars. Using hexadecimal characters and @ symbols.	**0.1**
Total weightage		**1.0**

Table 1 shows rules in the rule base and 10 different layers to evaluate website phishing criteria.

These rules which are used in the rule base divided into 10 layers. Each layer consists of two phishing characteristics rules andis assigned to 0.1 weightage if the rules in the any layer matches with website URL then it is given 0.1 score. The score of this websites is given from 0 to 1. Here 0 indicates low phishy website and 1 indicates high phishy website. The intermediate values between 0 to 1 indicates 'very legitimate' to very phishy sites.

The table below gives the different phishing intensity rating of the website for the scores obtained to the given inputs. Based on these results we can finally conclude whether website is fraudulent or the original one. Hence using fuzzy gives one of the efficient way to derive the website phishiness

Phishing intensity	**score**
highly Legitimate	0 – 0.1
Legitimate	0.1-0.3
suspicious	0.3 – 0.6
phishy	0.6 – 0.8
Highly phishy	0.8 – 1.0

Table 2 shows the intensity of phisiness from the given URL score.

V. CONCLUSION AND FUTURE WORK :

This fuzzy website phishing model detects the website phising using 10 layers which consists of rules. This techniques also shows the phishiness if some of the rules does not obey the phishing characteristics while come of the characteristics are clear. We can say that even some phishy characteristics are noticed it does not mean the entire website phishy. We can conclude website as phishy when it does not obey most of the characters.

As for future work we will propose and implement intelligent phishing detecting website using fuzzy logic and data mining algorithm techniques. So that we can able to detect the phishy websites in more advanced techniques and with more accurately**.**

ACKNOWLEDGMENT

We are pleased to express our sentiments of gratitude to all who rendered their valuable guidance to us. We express our appreciation and thanks to our college. We are also thankful to the our guide prof. DR B.BALAMURGAN

REFERENCES

[1] Shahriar, Hossain, and Mohammad Zulkernine. "PhishTester: automatic testing of phishing attacks." *Secure*

Software Integration and Reliability Improvement (SSIRI), 2010 Fourth International Conference on. IEEE, 2010.

[2] Zhou, Yu, et al. "Visual Similarity Based Anti-phishing with the Combination of Local and Global Features." *Trust, Security and Privacy in Computing and Communications (TrustCom), 2014 IEEE 13th International Conference on*. IEEE, 2014.

[3] Approximate string matching algorithm for anti phishing by Dona Abraham andNisha S Raj Department of Computer Science and Engineering SCMS School of Engineering and Technology

[4] Iacono, Luigi Lo, et al. "UI-Dressing to Detect Phishing." *High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC, CSS, ICESS), 2014 IEEE Intl Conf on*. IEEE, 2014.

[5] Yu, Weider D., ShrutiNargundkar, and NagapriyaTiruthani. "A phishing vulnerability analysis of web based systems." *Computers and Communications, 2008. ISCC 2008. IEEE Symposium on*. IEEE, 2008.

[6] Khan, Ahmad Alamgir. "Preventing phishing attacks using one time password and user machine identification." *arXiv preprint arXiv:1305.2704*(2013).

[7]. P.A. Barraclough, M.A. Hossain, M.A. Tahir b, G. Sexton, N. Aslam, " Intelligent phishing detection and protection scheme for online transactions", Expert Systems with Applications 40 (2013) 4697–4706.

[8]. V. Shreeram, M.Suban, P.Shanthi, K. Manjula, "AntiPhishing Detection Of Phishing Attacks Using Genetic Algorithm", ICCCCT'10. In proceeding of IEEE.

[9]. RadhaDamodaram, M.L.Valarmathi, "Phishing Website Detection and Optimization Using Particle Swarm Optimization Technique", International Journal of Computer Science and Security (IJCSS), Volume (5): Issue (5): 2011.

[10]. Gaurav Kumar Tak and Gaurav Ojha, "Multi-Level Parsing Based Approach against Phishing Attacks with the Help of Knowledge Bases", International Journal of Network security & its applications (IJNSA), Vol.5, No.6, November 2013.

[11]. Ammar Almomani , B. B. Gupta , Tat-chee Wan , AltyebAltaher , SelvakumarManickam, "Phishing Dynamic Evolving Neural Fuzzy Framework for Online Detection "Zero-day" Phishing Email", Indian Journal of Science and Technology, Vol: 6 Issue: 1 January 2013 ISSN:09746846.

[12]. Bryan Parno, Cynthia Kuo, and Adrian Perrig. "Phoolproof of Phishing Prevention", Financial Cryptography and Data Security, Springer, 2006