Proposed Approach to Detect Phishing Webpage Based on Multi-Browsers

Hala Bahjet Abdul Wahab and Thikra M. Abed*

Department of Computer Science - Technology University – Baghdad - Iraq

*thikra_mohamed@yahoo.com

Received: 24 September 2017 Accepted: 7 February 2018

Abstract

Phishing techniques have not just developed in number, but as well in sophistication. Phishers could have plenty of approaches and techniques to conduct a well-designed phishing attack. Developing countries such as Iraq may have been facing Internet threats like phishing. This paper aim to proposed efficient the phishing detection method. The proposed algorithm utilizes five information sources Google, Yahoo, Startpage, Bing and Mozilla then analyze the information retrieval from five search engines to characterize phishing site. The Several research techniques using single information source protection from Phishing which leads to inaccuracy in results. The proposed algorithm run online with combined between precision the analysis and strong the search engines. The experimental results which implement on 1000 URL demonstrate rate the true positive as 97.4% and the false positive rate 2.6% with true negative as 98.8% and false negative 1.2%.

Keywords: Anti-Phishing, Online Security, Phishing attacks, Browsers.

اقتراح نهج لكشف صفحات التصيد بالاعتماد على عدة متصفحات

هالة بهجت عبدالوهاب و ذكرى محمد عبد

قسم علوم الحاسوب - الجامعة التكنلوجية - بغداد

الخلاصة

لم تتطور تقنيات التصيد من حيث العدد فقط وانما من ناحية التطور العلمي والتكنلوجي ايضا .اذا ان المتصيدون يمتلكون الكثير من المناهج والتقنيات التي تكون مصممة بشكل جيد لاجراء هجوم التصيد . وقد تواجه البلدان النامية مثل العراق تهديدات على الانترنيت مثل التصيد الاحتيالي .تهدف هذه الورقة الى تصميم طريقة فعالة لكشف التصيد. الخوارزمية المقترحة تستخدم خمس مصادر للمعلومات وهي Mozilla، Yahoo Bing، Google وكذلك Startpag ثم يتم تحليل المعلومات المسترجعة من محركات البحث الخمسة لوصف الموقع المتصيد.العديد من التقنيات تستخدم مصدر واحد للمعلومة للحماية من التصيد الاحتيالي وهذا يؤدي الى عدم الدقة في النتائج. الخوارزمية المقترحة تنفذ ضمن الوقت الحقيقي وتجمع بين دقة التحليل وقوة محركات البحث .النتائج التجريبية التي نفذت على1000 رابط اوضحت بان نسبة الايجابية الصحيحة 97.4 % والايجابية الخاطئة بمقدار 2.6 % والسلبية الصحيحة 98.8 % والسلبية الخاطئة 1.2%.

الكلمات المفتاحية: مكافحة التصيد، أمن الأنترنيت، هجمات التصيد، المتصفحات.

Introduction

National threats to security contain those aimed against governmental systems and networks military as well as against private firms that support government actions or management important infrastructure threats to commerce including obtaining confidential intellectual property of private companies and governments, or persons with the goal of using this property for economic profit. unauthorized disclosure of personally identifiable data lead to threats the individuals such as taxpayer information, Social Safety numbers, credit and debit card data, or medicinal records. The revelation of such data could make hurt people included identity fraud cash loss, and humiliation. Famous Typical threats include the following:

• A bot-network operator which uses a network, distantly controlled systems to arrange assaults and to convey phishing plans, spam, and malware attacks.

• The criminal category which attacks systems of their goals monetary profit. In particular, sorted out criminal gatherings utilize phishing, malware/spyware, and online misrepresentation and spam to perpetrate wholesale fraud.

• International corporate spies which behavior industrial espionage and large-scale monetary steal and to hire or improve hacker skill.

• Phishers are people or little gatherings that implement phishing plans to take data for fiscal profit. Phishers should use spam and spyware or malware to complete their objectives.

• Spammers who are people or associations that convey spontaneous email with covered up or false data keeping in mind the end goal to offer items, lead phishing plans, appropriate spyware, or assault associations (e.g. a refusal of administration) [1].

Phishing

An easy Phishing is metaphorically similar to fishing in the water, but instead of trying to catch a fish, attackers try to steal consumer's personal information. When a user opens a fake webpage and enters the username and protected the password, the credentials of the user are acquired by the attacker which can be used for malicious purposes. Phishing websites look very similar in appearance to their corresponding legitimate websites to attract a large number of Internet users [2]. Phishing is an innovation-based, social building strategy where attackers endeavor to show up as approved sources. An expansion in online correspondence has the danger of phishing to such an extent that the accessibility and ubiquity of the Internet encourage cybercriminals' capacities to mount phishing assaults against various entities with one strike [3]. In like manner, the Anti-phishing working group announced no less than 277, 693 exceptional phishing websites worldwide in the fourth Quarter of 2016 [4].

Related Work

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

Recently researchers [5] had implemented a technique which is based on neuro fuzzy method. This strategy utilizes five information sources (pop up from sends, phish tank, client conduct plot, client determined destinations, advocated site rules) to characterize phishing site with

more exactness and precision in light of two overlay cross approval. Several research techniques are single folded protection which leads to 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.

The proposed method in [6] GoldPhish caught a picture of a page, at that point utilizes optical character recognition to change the picture to text, then use the Google PageRank algorithm to support render a decision on the truth of the site. In the wake of testing our tool on 100 real websites and 100 phishing websites, we carefully reported 100% of legitimate websites and 98% of phishing websites.

The researcher in [7] proposed PhishZoo is a phishing detection method uses the trusted websites profiles to detect attacks. This offers comparable precision to different techniques like blacklisted approach. The preferred standpoint in utilizing this approach is that it can classify different phishing methodologies and attacks on smaller websites (Intranets).

Proposal in [2] very useful approach for detecting phishing websites efficiently based on visual similarities techniques. Phishing website looks very similar in appearance to its corresponding legitimate website to trick users into confirming that they are browsing the correct website. Visual similarity based phishing detection techniques utilize the feature set like text content, text format, HTML tags, Cascading Style Sheet (CSS), image, and so forth, to make the decision. This approach compares the suspicious website with the corresponding legitimate website by using various features and if the similarity is exceeded than the predefined threshold value then it is declared phishing.

Spread of Phishing Attacks

Based on several types of research the main reasons following made users susceptible to the phishing attack and helped to spread the phishing as a threat to the web environment [8]:

1. Some users lack the basic knowledge of current online threats.
2. Unfamiliar threats make some users not be able to protect themselves although they perception how to deal with computer viruses and hackers they cannot generalize what they knew to unfamiliar threats.

3. Some users are careful of falling prey to phishing but they have not strategies to recognizing the phishing attacks and is usually the users focus on their main tasks, while they considered the security clues is a secondary task.
4. Some users may disregard the security indicators such as warning messages, while other users may notice these messages but they consider the warnings were invalid.
5. Some internet users have a shortage of knowing how the organizations present the online services and formally contacting for their consumers in case of maintenance and information update issues.

Methods to Combat Phishing

Generally, there are two methods to combat phishing:

1. Non-Technical Methods which consist of two domains, the first is legal solutions and the second is Education. The legal solutions aim to enact laws against phishing activities and in 2004 the phishing added to the computer crime list. The education solutions aim to consumer's education about importance notice the security indicators within the website. In this domain, many types of research published and recently, games appeared to train the users to deal with phishing attacks.
2. Technical Methods which appeared to deal with weaknesses in previously mentioned solutions. Several types of research studies, commercial and non-commercial solutions are offered to combat phishing. Awareness, education and using anti-phishing software applications are ways to defend against phishing attacks. Anti-phishing is providing a holistic approach towards the fight against phishing. It including technological innovation, legislation and law enforcement, industry collaboration, and consumer awareness [9]. Search Engine (SE) can used as technique to detect the phishing because this technique deal with assumption for detecting a web page as normal if web page appeared among the top search results obtained, as it is assumed that a normal web page would reach a higher index rank than a phishing web page which has remained active for a very short period of time. This assumption generates due to the fact that, phishing web page typically remain life over the internet for a very short period of time, thus making the probability of them being popular and indexed in SEs nearly impossible, Most SEs index websites after a

specific period of time, as the popularity or page rank and the number of hits increases with time, and most phishing websites remain up for a few that making indexing impossible, even if an index exists, it is very difficult for the phishing website to get a high rank in that short period of time [10].

Web pages Design

Scarcely there are properties and basic practices of web composition, for example:

1. Site brand names generally show up in specific parts of a page, for example, a title which renders the site identity accessible and recognizable.
2. The universal exercise of synchronizing the brand with a domain name loans authenticity to the technique of matching the textual brand name with domain keyword to locate if a domain real points to the site the brand indicate. The domain keyword is the section in the area speaking to the brand name, which is normally the non-nation character second-level area, for example, "Paypal" for "paypal.com".
3. Phishing web pages are much less likely to be crawled and indexed by major search engines than their legitimate counterparts due to their short-lived nature and few in-coming links.
4. Login form to demand important user data this behavior which usually phishing website gives, which alone could serve as a feature in classifying web pages [11].

In addition to that, the page after design will part of WWW. So, it has Page Rank which represents a numeric incentive on a size of [0, 1] that speaks to the relative significance of a page inside an arrangement of pages. The higher the Page Rank, the more vital the page. Phishing site pages are fleeting and, in this way, either have a low Page Rank or their Page Rank does not exist in the Crawl Database. Three-page rank highlights that give biased power are the Page Rank of URL, Page Rank of Host and whether the Page Rank is available in Crawl Database. All of the white rundown URLs have a host name Page Rank an incentive in the scope of [0.75, 1], demonstrating that host name Page Rank esteem is a solid element for distinguishing if a URL is non-phishing. Phishing site pages are typically open for just a brief timeframe; accordingly, many won't be found in the index [12].

Fuzzy Logic

Fuzzy logic utilizing from many decades in different topics like engineering and explores to insert the contributions to PC show for some applications, it is for the most part, helpful for individuals who include in innovative work and development. Fuzzy logic provides information to order the site phishing dangers. Compared with other Methodologies, the importance of fuzzy logic summarizes in use of the linguistic variables to describe the phishing degree. It allows the intermediate degree between notations like true and false, black and white, hot and cold, etc. as utilized in Boolean logic. In the fuzzy system, values are specified by numbers from zone 0 to 1 where zero appears absolute untruth and one appears absolute truthfulness. Fuzzy logic is utilized to evaluate the degree of phishing in a variety of web pages [13].

Fuzzy Inference is considered the main unit of a fuzzy logic the essential works making a decision, Figure (1) shows the basic components of a fuzzy inference system. It utilizes the "IF…THEN" rules forever with connectors "OR" or "AND" to build major decision rules. The output from Fuzzy Inference is always a fuzzy set irrespective of its input which can be fuzzy or crisp. The parallel If-Then rules in fuzzy inference form the deducing mechanism and refer how to imagine input variables onto output space. Generally, from knowing the Mamdani, Sugeno, and Tsukamoto are three types of fuzzy inference methods and the Mamdani is more popular [14].

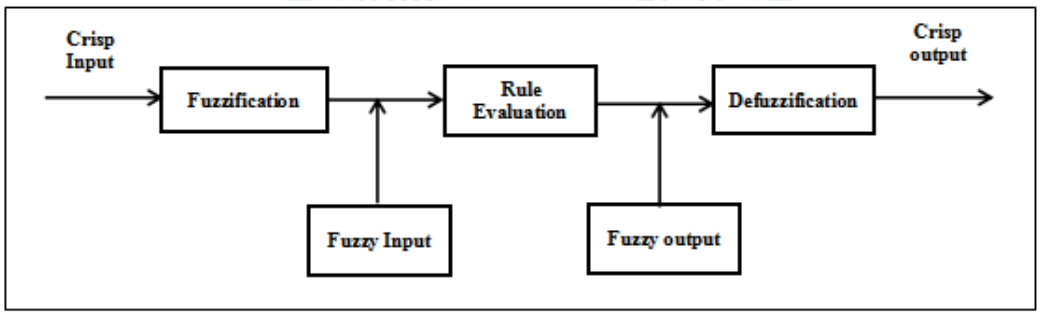

Figure 1: The basic components of a fuzzy inference system [15].

Mamdani Fuzzy Inference Methods

In 1975, Mamdani represented one of the first fuzzy systems which applied a set of fuzzy rules supplied by experienced human operators to control a steam engine and boiler combination. This approach has been successfully used in a variety of industrial problems. Mamdani inference process is an aggregation of the resulting composition operations for each rule, it should be mentioned that aggregation could be done by different operators, such as arithmetic, geometric or harmonic means, MAX and MIN. In the defuzzification step, the output fuzzy number of each rule is explained by the composition between a fuzzy singleton and the implication relation, output fuzzy numbers are changed to a crisp number [15].

Proposed Model

Real-time protection of phishing attacks on legitimate web pages is our goal in this paper. To enable this, we proposed an algorithm to test a link by information retrieval about this link which found in Google, Yahoo, Bing, start page and Mozilla search engines based on the application programming interfaces (APIs). Then analysis this information to indicate whether the URL is legitimate or phishing.

Anti-Phishing proposed to protect users from zero-day phishing attacks, it utilizes a fact; the trusted URL for legitimate organizations in most have the brand name for organizations as domain name or as site title and generally, the phishers when attack a legitimate site are impersonating name of the domain close up to the name of legitimate domain and web page title similar to the legitimate site title to make users believe this authorized site. Traditionally, phishing attacks targeted at the interactive web pages such as bank web page and Paypal web page which require personal information. In the following figure (2) shown the flowchart of the proposed algorithm and its explanation.

Proposed Approach to Detect Phishing Webpage Based on Multi-Browsers

Hala Bahjet Abdul Wahaband Thikra M. Abed

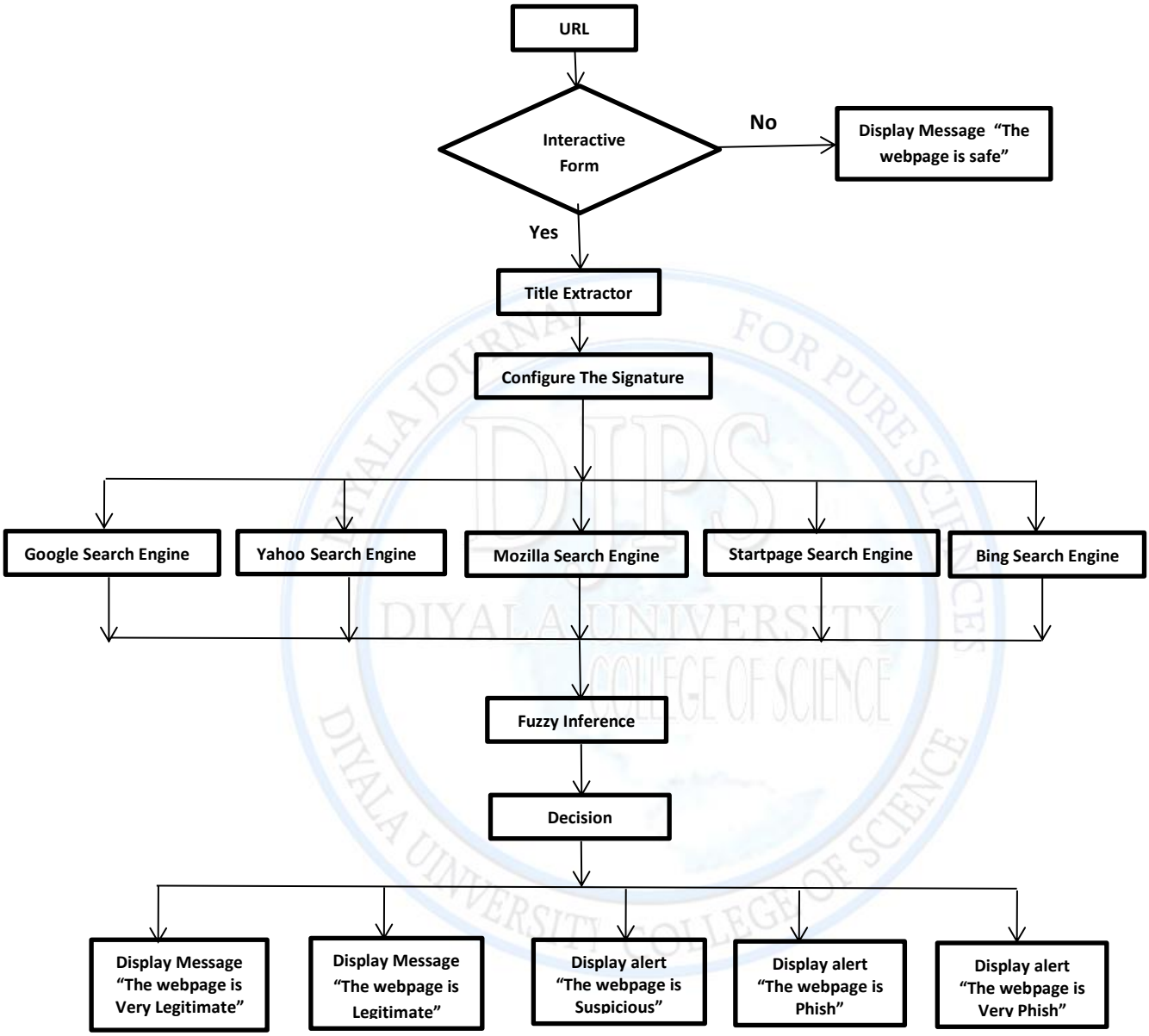

Figure 2: Flowchart of Proposed Algorithm

The Proposed Algorithm:

Input: User enters a link to test if it legitimate or Phish.

Output: The decision as a message to the user legitimate or Phish.

Processing:

{Phase 1: Interactive Form Detector}

Step-1: URL opened as the HTML source code.

Step-2: Search for specific features in HTML source code.

Step-3: If step-2 find an interactive feature that means the URL have phishing indicator and go to the second phase but if HTML source code does not have interactive features a message display to user "the web page is safe".

{Phase 2: Extract Host Names Part}

Step-4: Extract web page title from HTML source code.

Step-5: Configure the signature by merge URL with web page title.

Step-6: Send the signature as the request to five API search engines (Google, Yahoo, Mozilla, Bing, and start page) to bring the top 10 results.

Step-7: Analysis the HTML source code for a response each search engine to extract the top 10 URL.

Step-8: Extract Host name part of all URLs and Host name part of URL the user wants to check.

{Phase 3: The Decision}

Step-9: Compare the host name for URL the user wants to test with each one of ten hast name for Google search engine in order to vote high or low.

Step-10: Compare the host name for URL the user wants to test with each one of ten hast name for Yahoo search engine in order to vote high or low.

Step-11: Compare the host name for URL the user wants to test with each one of ten hast name for Mozilla search engine in order to vote high or low.

Step-12: Compare the host name for URL the user wants to test with each one of ten hast name for Bing search engine in order to vote high or low.

Step-13: Compare the host name for URL the user wants to test with each one of ten hast name for startpage search engine in order to vote high or low.

Step14: The output from step-9 to step-13 enter to Fuzzy Inference to take the decision.

Step-15: Display the decision as a message to the user.

End.

A. Test Dataset

Our proposed tested on 1000 URL, 500 phishing link are download from PhishTank site and 500 legitimate links are download from Alexa site. Without eliminating the links which back to the web pages that written in non-English text. We download databases from monthly report that available on phishtank.com then manually examined these links to ensure the phish web pages are active (valid) online to save HTML source code for them that back to the fact the phish web page may close in any time and that makes us unable to measure the efficiency of the Anti-phish proposed because our proposed based on test the HTML source code for URL. With a legitimate website, we downloaded a list of million sites that available of alexa.com and take top 500 sites.

B. The First Phase

The first phase is allowed to enter the link to test. This URL opened as the HTML source code then filtering HTML source code by interactive form detector to extract the content features we consider them as indicators for the interaction with the user. Hence, The second phase is implemented if HTML source code contains any indicator for the interactive but if non-contain, the proposed consider the link is legitimate and display a message to the user to ensure the link is safe. This phase aims to check if a web page has indicators for phishing attacks.

Interactive Form Detector

In this section the heuristics-based algorithm presented to distinguish the interactive web pages, it deals with HTML tags and java scripting language. The Interactive Form Detector define set of features which consider as indicators for interactive web pages, the features are [FORM tags such as submit and button tag, post and get method tags, javascript tags, and popup boxes tags such as confirm box tags, prompt box tags and alert box tags, paragraph tags]. Usually, such these features make users interact with web pages and enter the personal information e.g. password or answer to questions found in a web page and that makes the phisher able to analyze these answers and get on important information about users. Due to

the phishing and other suspicious behaviors, we proposed the following steps to advertise a presence in an interactive form:

Step-1: search on the javascript tags, in web design JavaScript Language, is a must-have interactive component. These tags searched in the source code for URL wanted to checks and the search return true if it found javascript tags then convert web page to the second phase. The reason behind this step is to protect the user from any form visible or hiding took an information from the user then analysis it to get the private information and because the JavaScript language allows interacting with users that make our approach take precautions from web pages that carry JavaScript tags.

Step-2: search on other features in HTML source code refer to a suspicious web page. From these features the paragraph Tags, this step search on text between which represent text display in the web page, if the search does not find text that refers to use only images to build the web page and this style used by the phishers to avoid detection by text similarity checked and this step also includes check if found: form tag, post and get method tag and any type of pop up boxes tags (prompt, alert, confirm).

Step-3: if two above steps implement and return false that means the web page does not have interactive features and it only displays information but if return true that convert HTML source code to the second phase.

Used the interactive detector helps to take fast decisions for web pages which does not contain suspicious interaction, thus accelerating the detection process. We test this algorithm on our phish database and all web page phishing recognize as interactive web pages.

C. The Second Phase

The second phase is applied to the URL which has the phishing indicators, in this phase the title extract of HTML source code. Title in a web page is important because usually a page title contains the brand name of the site, here extracted the web page title by use pattern matching with title tag then build signature contains the URL and the web page title. The signature sends a request to the five search engines in the same time using API Google, API Yahoo, API Start page, API Bing, and API Mozilla to retrieval the top 10 results from each search engine. The second Phase receives the top 10 results as HTML source code then

analysis it to extract the top 10 URL. The process extract URLs from HTML source code depend on our study for the source code of response each search engine of five search engine and we could determine the tags which carrier these URLs, so we use the pattern matching for specific tags for each search engines to extract the top 10 URL and saved in the array. After that, this phase extract only host name part from each URL extracted from each search engine and extract host name part from link which enters to test. However, the output from this phase array contains 10 host name for each search engines in addition to the host name for URL the user wants to test. After our study of problems in anti-phishing existing which depend on the search engine, we proposed merge title with URL in signature to reduce the false positive which happen because the newly legitimate site launch to the web have not high rank and to bring web pages which written in non-English text to the top results.

Request from Search Engines

Signature passes as a request to the five API search engines (Google, Yahoo, Mozilla, Bing, and Start page). We take the response from each search engine as HTML source code and analysis this source code to extract top 10 URL. Then extract Host name part of all URLs and Host name part of URL the user wants to check.

D. The Third Phase

The final phase contain with comparing the host name for URL the user wants to test with each one of ten hast name for Google search engine (first voter) in order to vote, if happen to match the first vote as "high" but if non-match happening the first vote as "low", this procedure also implemented with hast names for Yahoo, Startpage, Bing and Mozilla after all comparisons the third phase has five voices. This phase classifier the voices using Fuzzy Inference. Fuzzy inference utilizing a set of rules to determine the level of phishing risk in the link and display the decision a message to the user. The proposed implement online and taken the decision in 22 seconds.

Fuzzy Inference Process

We proposed using the Mamdani fuzzy inference to classify the five voices in order to reduce the false positive so the proposed implement the following five steps:

First Step: Fuzzify The Input Variables

The crisp values of input variables are transformed into the equivalent membership values. Hence, there are five input variables describe response the search engines with crisp value "low" if response the search engine non-matching with host name part for the URL is entered and "high" if matching, the equivalent membership values will be 0 for low and 0.2 for high.

Second Step: Apply the Fuzzy Operator

AND operation and OR operation are the most common fuzzy operators. We use AND operation.

Third Step: Apply the Implication Method

Generally, the form a single fuzzy If-Then rule as follows:

If x is A, then y is B

The antecedent is the first If-part, where x is input variable and the rest Then-part is called the consequent, and y is output variable. When we build If-Then rules suggested the antecedent as two weight 0 and 0.2 which represent low and high for the five voices which produced from the response of the search engines and the consequent part of If-Then rule are fuzzy sets within the interval between 0 and 1(five intervals). The output represents website risk assessment, so the rule base contains $(2^5) = 32$ entries. The structure rules base is shown in figure (3) and all rules base in table1.

Proposed Approach to Detect Phishing Webpage Based on Multi-Browsers

Hala Bahjet Abdul Wahaband Thikra M. Abed

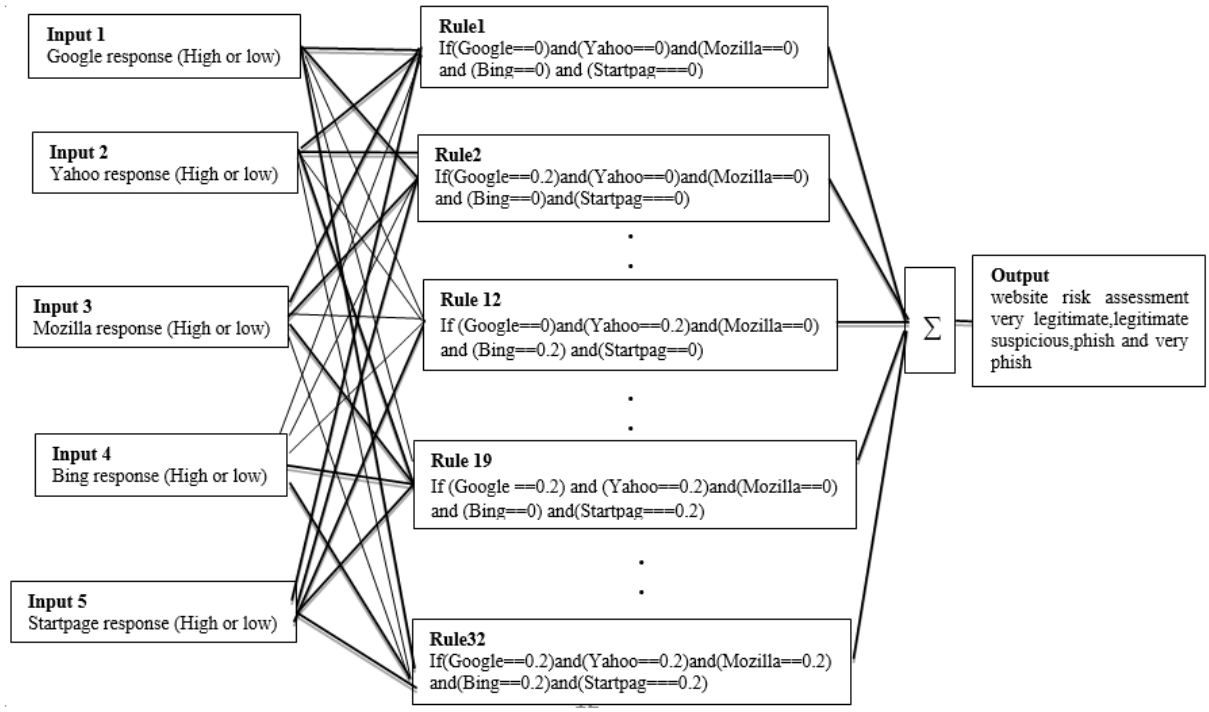

Figure 3: The Structure of Rules Base.

Table 1: The Rules base

No.rule	Rules
1	If (Google ==0) and (Yahoo==0)and (Mozilla==0) and(Bing==0) and (Startpag==0) Then output==0
2	If (Google == 0.2) and (Yahoo==0)and (Mozilla==0) and (Bing==0) and (Startpag==0) Then
3	If (Google ==0) and (Yahoo==0.2)and (Mozilla==0) and (Bing==0) and(Startpag==0) then
4	If (Google ==0) and (Yahoo==0)and (Mozilla==0.2) and (Bing==0) and (Startpag==0) Then
5	If (Google ==0) and (Yahoo==0)and (Mozilla==0) and (Bing==0.2) and(Startpag==0) Then
6	If (Google ==0) and (Yahoo==0)and (Mozilla==0) and (Bing==0) and(Startpag==0.2) Then
7	If (Google ==0.2) and (Yahoo==0.2)and (Mozilla==0) and (Bing==0) and(Startpag==0) then
8	If (Google ==0.2) and (Yahoo==0)and (Mozilla==0.2) and (Bing==0) and(Startpag==0) Then
9	If (Google == 0.2) and (Yahoo==0)and (Mozilla==0) and (Bing==0.2) and(Startpag==0) Then
10	If (Google ==0.2) and (Yahoo==0)and (Mozilla==0) and (Bing==0) and(Startpag==0.2) Then
11	If (Google == 0) and (Yahoo==0.2)and (Mozilla==0.2) and (Bing==0) and(Startpag==0) Then
12	If (Google ==0) and (Yahoo==0.2)and (Mozilla==0) and (Bing==0.2) and(Startpag==0) then
13	If (Google == 0) and (Yahoo==0.2)and (Mozilla==0) and (Bing==0) and(Startpag==0.2) Then
14	If (Google ==0) and (Yahoo==0)and (Mozilla==0.2) and (Bing==0.2) and(Startpag==0) Then
15	If (Google == 0) and (Yahoo==0)and (Mozilla==0.2) and (Bing==0) and(Startpag==0.2) Then
16	If (Google ==0) and (Yahoo==0)and (Mozilla==0) and (Bing==0.2) and(Startpag==0.2) Then
17	If (Google ==0.2) and (Yahoo==0.2)and (Mozilla==0.2) and (Bing==0) and(Startpag==0)Then

18	If (Google ==0.2) and (Yahoo==0.2)and (Mozilla==0) and (Bing==0.2) and(Startpag==0) Then
19	If (Google ==0.2) and (Yahoo==0.2)and (Mozilla==0) and (Bing==0) and(Startpag==0.2) Then
20	If (Google ==0.2) and (Yahoo==0)and (Mozilla==0.2) and (Bing==0.2) and(Startpag==0) Then
21	If (Google ==0.2) and (Yahoo==0)and (Mozilla==0.2) and (Bing==0) and(Startpag==0.2) Then
22	If (Google ==0.2) and (Yahoo==0)and (Mozilla==0) and (Bing==0.2) and(Startpag==0.2) Then
23	If (Google ==0) and (Yahoo==0.2)and (Mozilla==0.2) and (Bing==0.2) and(Startpag==0) Then
24	If (Google ==0) and (Yahoo==0.2)and (Mozilla==0) and (Bing==0.2) and(Startpag==0.2) then
25	If (Google ==0) and (Yahoo==0)and (Mozilla==0.2) and (Bing==0.2) and(Startpag==0.2) Then
26	If (Google ==0) and (Yahoo==0.2)and (Mozilla==0.2) and (Bing==0) and(Startpag==0.2) Then
27	If (Google ==0) and(Yahoo==0.2)and (Mozilla==0.2)and(Bing==0.2) and(Startpag==0.2) Then
28	If (Google ==0.2) and(Yahoo==0)and (Mozilla==0.2)and (Bing==0.2)and(Startpag==0.2) Then
29	If (Google ==0.2)and (Yahoo==0.2)and(Mozilla==0) and (Bing==0.2)and(Startpag==0.2) Then
30	If (Google ==0.2) and (Yahoo==0.2)and (Mozilla==0.2)and(Bing==0)and(Startpag==0.2) Then
31	If (Google ==0.2) and (Yahoo==0.2)and (Mozilla==0.2)and(Bing==0.2)and(Startpag==0) Then
32	If (Google ==0.2)and (Yahoo==0.2)and(Mozilla==0.2)and(Bing==0.2)and(Startpag==0.2) Then output==1

Four Step: Apply The Aggregation Method

After generating the If-Then rules, the aggregation process is implemented to combine the input weights, the result represent the final decision in scalar quantity.

Five Step: Defuzzification

After aggregation all responses of search engines (see table2) the output will be a single scalar quantity represents the final membership values and in order implement the defuzzification process we defined the membership values for each linguistic values as shown in table3.

Table 2: Total weight each response.

Criteria	Weight
Google response	0.2
Yahoo response	0.2
Mozilla response	0.2
Bing response	0.2
Startpag response	0.2
Total Weight	1

Table 3: Linguistic values with membership values.

Linguistic Values	Membership Values
Very phish	[0-0.2)
phish	[0.2-0.4)
Suspicion	[0.4-0.6)
legitimate	[0.6-0.8]
Very legitimate	(0.8-1]

Results

The results build from test the phish and legitimate dataset. Usually, any system deal with phishing attacks produces four rates as a measure of his performance: **true positive(*tp*)** which mean the system determine legitimate link as legitimate web page, **false positive(*fp*)** which mean the system determines legitimate link as phish web page, **true negative(*tn*)** which mean the system determine phish link as phish web page and **false negative(*fn*)** which mean the system determine phish link as legitimate web page. When we test the proposed anti-phishing tool for 1000 links, the results of 500 Phish links showed in table 4 and the results of 500 legitimate links showed in table5.

Table 4: The Results of the Phish Dataset.

	Of 500 Phish link, the number links which classify as	The rates	Anti-phishing tool performance in phish dataset as:	
Very-legitimate webpage	*0*	0%	False Negative	1.2%
Legitimate web page	*6*	1.2%		
Suspicious web page	*36*	7.2%	True Negative	98.8%
Phish web page	*38*	7.6%		
Very phish web page	*420*	84%		

Table 5: The Results of the Legitimate Dataset.

	Of 500 Legitimate link, the number links which classify as	The rates	Anti-phishing tool performance in the Legitimate dataset as:	
Very-legitimate web page	405	81%	True Positive	97.4%
Legitimate web page	82	16.4%		
Suspicious web page	4	0.8%	False Positive	2.6%
Phish web page	2	0.4%		
Very phish web page	7	1.4%		

The *ex*perimental results *demonstrate highly efficient performance* for *the* Anti-phishing proposed. The Confusion Matrix for our proposed reach to rates showed in table 6 which our enable from compute the down metrics.

Table 6: Matrix Confusion for Two Dataset.

	actually Positive Class	actually Negative Class
Predict Positive Class	487 (True Positive)	6 (False Negative)
Predict Negative Class	13 (False Positive)	494 (True Negative)

Accuracy (acc) [16] = $\frac{tp+tn}{tp+fp+tn+fn} = \frac{487+494}{487+13+494+6} = 98.1\%$

Error Rate (err) [16] = $\frac{fp+fn}{tp+fp+tn+fn} = \frac{13+6}{487+13+494+6} = 1.9\%$

Sensitivity (sn) [16] = $\frac{tp}{tp+fn} = \frac{487}{487+6} = 98.78\%$

Specificity (sp)[16] = $\frac{tn}{tn+fp} = \frac{494}{494+13} = 97.44\%$

Precision (p) [16]= $\frac{tp}{tp+fp} = \frac{487}{487+13} = 97.4\%$

Conclusion

Phishing became a serious web threat leads to lack of confidence for the e-commerce. The proposed algorithm succeeds to detect the phishing links. Using the information retrieval from five search engines and analysis processes to contents of source code, and using Fuzzy inference give efficient results to reach a final decision. The proposed technique able to take true decisions with zero-day phishing links and not only deal with newly web pages but also web pages which are written in non-English text, it considers adaptive anti-phishing tool based on dynamic sources and not requiring existing training data. The proposed technique can be used as an intelligent browser extension.

References

1. Clark, Robert M., and Simon Hakim. "Protecting Critical Infrastructure at the State, Provincial, and Local Level: Issues in Cyber-Physical Security." Cyber-Physical Security. Springer International Publishing, 2017; pp 1-17.
2. Jain, Ankit Kumar, and B. B. Gupta. "Phishing Detection: Analysis of Visual Similarity Based Approaches." Security and Communication Networks 2017 (2017).
3. Furnell, Steven. "It's a jungle out there: Predators, prey and protection in the online wilderness." Computer Fraud & Security 2008.10 (2008): 3-6.
4. Anti-Phishing Working Group. (2017). Phishing activity trends report. Anti-Phishing Working Group.
5. Barraclough, P. A., et al. "Intelligent phishing detection and protection scheme for online transactions." Expert Systems with Applications 40.11 (2013): 4697-4706.
6. Dunlop, Matthew, Stephen Groat, and David Shelly. "Goldphish: Using images for content-based phishing analysis." Internet Monitoring and Protection (ICIMP), 2010 Fifth International Conference on. IEEE, 2010.
7. Almomani, Ammar, et al. "Phishing dynamic evolving neural fuzzy framework for online detection zero-day phishing email." arXiv preprint arXiv:1302.0629 (2013).
8. Mohammad, Rami M., Fadi Thabtah, and Lee McCluskey. "Tutorial and critical analysis of phishing websites methods." computer science review 17 (2015): 1-24.

9. Abdelhamid, Neda, Aladdin Ayesh, and Fadi Thabtah. "Phishing detection based associative classification data mining." Expert Systems with Applications 41.13 (2014): 5948-5959.

10. Varshney, Gaurav, Manoj Misra, and Pradeep K. Atrey. "Improving the accuracy of Search Engine based anti-phishing solutions using lightweight features." Internet Technology and Secured Transactions (ICITST), 2016 11th International Conference for. IEEE, 2016.

11. Xiang, Guang, and Jason I. Hong. "A hybrid phish detection approach by identity discovery and keywords retrieval." Proceedings of the 18th international conference on World wide web. ACM, 2009.

12. Garera, Sujata, et al. "A framework for detection and measurement of phishing attacks." Proceedings of the 2007 ACM workshop on Recurring malcode. ACM, 2007.

13. K. N. Manoj Kumar & K. Alekhya "Detecting Phishing Websites using Fuzzy Logic." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET).2016.

14. Wang, Chonghua. A study of membership functions on mamdani-type fuzzy inference system for industrial decision-making. Lehigh University, 2015.

15. Pourjavad, Ehsan, and Rene V. Mayorga. "A comparative study and measuring performance of manufacturing systems with Mamdani fuzzy inference system." Journal of Intelligent Manufacturing (2017): 1-13.

16. Hossin, M., and M. N. Sulaiman. "A review on evaluation metrics for data classification evaluations." International Journal of Data Mining & Knowledge Management Process 5.2 (2015): 1.