Detect and Prevent Phishing based on Hybrid Approach

Hala Bahjet Abdul Wahab[*]
Ph.D. (Asst Prof.)

Thikra M. Abed*
M.Sc.(Asst.Lec.)

Abstract: Phishing is one of the web threats, A phishing attack is considered successful if the user unknowing submits his personal information to the attacker such these attacks become popular in recent times. Developing countries such as Iraq may have been facing internet threats like phishing. This paper aim to build system based on the proposed algorithm to detect and prevent phishing attacks, the proposed analysis the structure and the HTML source code for URL with depend on the Google and Yahoo search engines in detect operation. The prevent operation based on proposed a method of secret sharing to product two shares, one for the user and other for the server with private keys for them. The experimental results which implemented on 300 URL demonstrated rate the true positive at 99.4% and the false positive rate 0.6% with true negative as 98% and false negative 2% with compare the false results with reports which released from the Google web safe and PhishTank in addition to build non-expanded secure shares in prevent operation with able to reconstruct same quality original image.

Keywords: Detect the Phishing, Cybercrime, Secret sharing, Browser extensions.

[*] University of Technology

1. Introduction

Internet is important to organizations which doing business on-line such as online trading, but the web-threats may be vulnerable the individual users to loss of private information, identity theft, financial damages, and loss of customer's confidence in e-business. Phishing is defined as the art of impersonating a website of an honest enterprise aiming to acquire private information, phishing attacks start with combine the social engineering and the technical tricks, not only increased number of phishing websites make such these attacks serious problem but also the smart strategies which used to design such websites [1] . In the early 1990s, with the growing popularity of the Internet, a new type of cybercrime birth;that is phishing, and the first recorded attack was in 1995 [2] . In February 2017, Anti-Phishing Working Group (APWG) releases a report about phishing trends, show 1,220,523 phishing attacks in 2016 was happening that means a 65% over 2015 and in the final quarter of 2004 the APWG saw 1,609 phishing attacks for every month, while in the final quarter of 2016 the APWG saw a medium of 92,564 phishing attacks for each month, an expansion of 5,753% over 12 years [3]. From application used to validate from phishing links are phishtank and Google Safe Browsing. Google Safe Browsing and phishtank enables client applications to validate whether a given URL contain phish attack or not [4]

2. Related Work

A literature review was performed to cover some of the related works and to provide an outline of the previous vital works in detecting phishing schemes some of them are:

M. Dunlop [5] Proposed Gold Phish technique which uses the optical character recognition on an image captured of a web page to convert it to text, then takes a decision about the validity of the site based on the Google PageRank algorithm. Effectiveness of Gold Phish is limited by the textual content available on a page, style of text, logos, fonts used and accuracy the OCR image.

Ashwaq T. et al. [6] presented the secret sharing to generate the secret shares based on wavelet transform in the case of the colored image, permutation key with time stamp used to scrambling the transparencies with wavelet coefficient is too hard to be tracked and the secure shares which generated do not reveal any information until all shares combine in specific way with existence the private key.

RB. Basnet et al. [7] discussed a study of anatomy a number of publicly available features on URLs alone with give a compare performance of different machine learning techniques. The aim of the study is able to anatomy of phishing URLs which created to trick users into divulging personal data. Al-Khalid et al. [8] presented technique to encrypt the color image based on a private key and exploiting the human vision system. The two shares are generated which were the same size as the secret color image and the experimental results showed better level of security and value of PSNR with less time of computation.

3. Search Engine as Detect Phishing Techniques

Search Engine (SE) can used as a technique to detect the phishing web pages based on assumption the normal web page could appear among a higher index rank than a phishing web page, this assumption is generated due to the fact that the phishing web page typically remained live for a very short period of time that's make indexed there in SEs nearly impossible because the page rank and the number of hits increases with time and most the search engines are indexing websites after a specific period of time [9]. GoldPhish and other schema which works on Google Search API implementation only on entering English-readable text. With Google’s limitation which require only 50 query term some of previous research is submitted to the text as line-by-line. When the Google Search API implemented is return by default the first ten results. In order to the legitimate website have high Page Rank it’s come up within the first ten results in a search engine, but newly webpage which launched over the web will consider as phishing web pages because their low ranking [5] .

3.1 A Uniform Resource Locator

A Uniform Resource Locator (URL) is created to address web pages. The URL begins with a protocol used to access the page. The phisher can change Free URL at any time to create a new URL. Unique part of the web site is “Domain Name” and the attacker must intelligently choose the domain names because the aim should be convincing the users and then setting the Free URL to make detection difficult [4] .

3.2 Majority Vote

Majority Vote is one of classification types which depend on consensus is proposed by Rahman et al. in 2002 which states if n independent experts produces a unique decision regarding the identity of the unknown sample and the experts have the same probability of being correct, then the sample is assigned to the class for which there is a consensus [10] . This principle employs in the decision phase to assign if the URL is a phish or not.

3.3 Chaotic Map

Chaos theory describes the behavior of certain nonlinear dynamic system that under specific conditions exhibit dynamics that are sensitive to initial conditions. Chaotic map is used to produce the chaotic sequence and used to control the encryption process. The chaos streams are generated by using various chaotic maps [11] :

3.3.1 The Logistic Chaotic Map

In most the classical one-dimensional (1D) logistic map used in image encryption algorithms due to not only it has good chaoticity with desirable autocorrelation and cross-correlation properties but also simple structure with only one control parameter and one initial condition. However, an intertwining logistic map is designed with more control parameters and more initial conditions to overcoming about some weaknesses which generates weak keys as pointed out and analyzed by such as blank windows, stable windows, uneven distributions of iterated sequences, etc. [12].

3.3.2 Piecewise Linear Chaotic Map

The skew tent map belongs to a category of the simple chaotic system and also its widely basic one in the piecewise linear chaotic map (PWLCM). The discrete version of the skew tent map is as follows:

$$x_{n+1} = f(x_n) = \begin{cases} x_n / a, & 0 \langle x_n \langle 0.5, \\ (1 - x_n)/(1 - a), & 0.5 \leq x_n \langle 1, \end{cases} \tag{1}$$

The system goes into a chaotic state when the parameter a locate in the open interval (0, 1). Thus, this system is an iterative sequence with the feature of strong randomness and unpredictability in addition to its sensitivity to initial value x_0. Hence, it often uses to encrypt the plaintext and the encryption process is also reversible. The basic piecewise linear map has the following expression:

$$f_p(x) = \begin{cases} x/p, & 0\langle x\langle p, \\ (x-p)/(0.5-p), & p \le x\langle 0.5, \\ (1-x-p)/(0.5-p), & 0.5 \le x\langle 1-p \\ (1-x)/p, & 1-p \le x\langle 1, \end{cases} \qquad (2)$$

Given the initial value x_0 and the control parameter p, where the range of the control parameter p is p ∊ (0, 0.5) we can get a random sequence {f_p(xi)} has some good chaotic properties such as strong randomness and unpredictability, etc., which every value is between 0 and 1 [13] .

3.3.3 Hyperchaotic System

The hyperchaotic system refers to as a family of chaotic systems used as candidates for encrypting the multimedia data which have more than one positive Lyapunov exponent [at least four] and have high efficiency, high security and high-capacity [14] . The hyperchaotic system generates the pseudorandom sequence with Chen's hyper-chaotic system which described in eq.(3). In eq. (3), the system parameters control are a, b, c, d, k when a = 36, b = 3, c = 28, d = 16 and k is set at 0.2 with initial conditions (0.3,-0.4,1.2,1) of t_1,t_2,t_3,and t_4 which can consider secret key[15]

$$\left.\begin{array}{l} t_1=a\,(t_2 - t_1)\ ; \\ t2= -\ t1t3 + d\ t1 + ct2 - t4\ ; \\ t3= t1t2 - bt3\ ; \\ t_4=t_1 + k; \end{array}\right\} \qquad (3)$$

3.4 Message Digest Algorithm

Message Digest Algorithm 5 (MD5) is hash a function processing and very fast to break up the input message into blocks of 512-bit and operates on a 128-bit state. MD5 is composed of four analogous rounds (sixteen, 32-bit words) with three types of operations: bit-wise Boolean, modular addition, and cyclic shift operations [16] .

3.5 Deoxyribo nucleic Acid Sequence

Adenine (A), Thymine (T), Cytosine (C) and Guanine (G) are four types of nucleotides which existing in deoxyribo nucleic acid (DNA) which defined as a type of molecules. A, C, G and T in DNA sequence used to encode 00, 01, 10 and 11. A and T, G and C are complementary just because in the binary system 0 and 1 are complementary, the DNA match rules are listed in following Table (1) [17] . DNA cryptography is defines as an implementation tool used an information carrier and modern biological technology base on DNA rules and DNA operation such as addition, subtraction and XOR, it appears as a new cryptographic area [18] .

Table 1: of DNA [17]. **Map rule sequence**

	1	2	3	4	5	6	7	8
A	00	00	01	01	10	10	11	11
C	10	01	11	00	11	00	10	01
G	01	10	00	11	00	11	01	10
T	11	11	10	10	01	01	00	00

3.6 Secret Sharing

A Secret image sharing (SIS) is splitting the secret image into noise-like n shadow images (also called shares or shadows), the secret image can be reconstructed by shadow images. SIS can be applied in information hiding, access control, watermarking, data security, authentication, and transmitting passwords etc. [19] . Secret Image Sharing and Visual cryptography (Special case of SIS) are tasks to protect image privacy against unauthorized data access and from could with a visual cryptographic scheme for color images used a secret key to encrypt the image and the division of encrypted image is done using Random Number, without the secret key the original image cannot be decrypted [6]. Most Visual Cryptography schemes (VCS) are designed to avert cheating attacks usually all VCS have drawbacks include the following: the scheme needs an online trusted authority, or it requires additional shares for the purpose of verification, or it has to sacrifice the properties by means of pixel expansion and contrast reduction of the original VCS [20] .

4. The Proposed Algorithm

The proposed algorithm protects users of the phishing attacks based on detect and prevent phishing URL. The Proposed consist of three phases: the first phase based on analysis the structure of a URL and navigation from through Google API and Yahoo API search engines, the second phase takes the decision with suggested alternative sites if the URL is phish or convert user to the third phase if the URL is legitimate. The third phase prevents the false negative from access to user and steal sensitive information's based on proposed a method of secret sharing. The block diagram of the proposed system has been shown in figure (1).

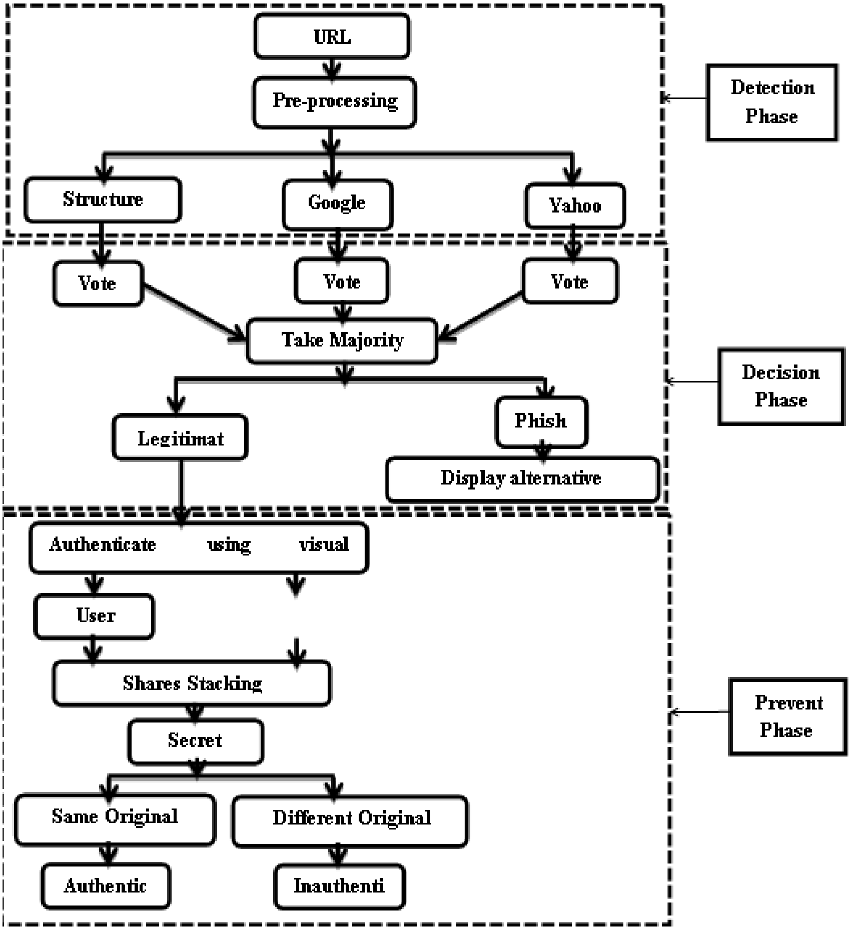

Figure 1. Block Diagram of Proposed Algorithm.

The Proposed Algorithm

Input: User enters a URL.
Output: Display the decision as a message to the user with suggested alternative sites if the URL is a phish or open URL if legitimate to perform the authentication operation between user and server.
Processing:

{Detect Phase}

Step-1: Check the structure of a URL entered if has two of phishing features indicators this step vote as (0) or vote as (1) if URL structure non-contain any phishing feature.
Step-2: URL opened as the HTML source code.
Step-3: Extract web page title from HTML source code.
Step-4: Configure the signature by merging the URL entered with the web page title.
Step-5: Send the signature as the request to Google search engine API and Yahoo search engine API to bring the top 30 high rank.
Step-6: Analysis the HTML source code the response Google search engine to extract the top 30 URLs then extract only host name part from each URL.
Step-7: Analysis the HTML source code the response Yahoo search engine to extract the top 30 URLs then extract only host name part from each URL.
Step-8: Extract host name part of the URL entered, then compare it with each one of 30 host names for Google search engine in order to vote 1 if happening matching or 0 in case of non-matching happened and compare it with each one of 30 host names for Yahoo search engine in order to vote 1 if matching or 0 if non-matching.

{DecisionPhase}

Step9: Take the decision about the web page from through the Majority vote.
Step10: if the URLis Phish suggest sites from top results the Google and Yahoo engines as alternative sites to the user but if legitimate Go to Step 11.

{Prevent Phase}

Step11: User enter private user key and loaded his share with original image and the server have time stamp and server share and private server key which represent the initial values for hyperchaotic, after that the legitimate server able to rebuild the secret color image from two shares then display it to the user, if same the secret image which user is choosed in register phase the web page is legitimate and user can enter’s his sensitive information otherwise is phish and user abstain present his information.

4.1 Dataset

Our proposed tested on 300 URL, 150 phishing link are downloaded from the PhishTank site (which it is considered one of the primary phishing-report [21] and150 legitimate link are downloaded from Alexa site. Without eliminating the links which back to the web pages that written in non-English text. We download phish databases 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 closed at any time and that makes us unable to measure the efficiency of the detect phish proposed because we proposed based on the HTML source code for URL in addition to structure of a URL. With legitimate website, we downloaded list of the million sites on alexa.com and take top 150 sites.

4.2 Detect Phase

This phase includes three parts, the first part is analysis the URL to extract the phishing features indicators based on rules, the second is bringing top results of Google search engine for the signature, which build in preprocessing from merge URL and title web page and the third part is bringing top results for the signature from Yahoo search engine, these parts explain as follow.

4.2.1 Analysis Structure of URL

To analysis the features which found in structure of a URL which consider as a first voter, nine features extract from the URL entered depends on rules take of [7] and feature-9 proposed as follows:

Feature-1: Search if IP address represents the URL domain name.

Rule: If IP address exist in domain, then suspicious URL else legitimate URL.

Feature-2: Ensure from average the URL length if number URL characters are equal to 54 or greater than 54.

Rule: If URL length < 54 then its legitimate URL else is suspicious URL.

Feature-3: When using @ symbol in the URL the browser considers the part after @ is real address and ignored the previous part.

Rule: If the URL has @ symbol, then suspicious URL else legitimate URL.

Feature-4: When using // symbol in the URL the user may directed to another site.

Rule: If more than one // symbol in URL then suspicious URL else legitimate URL.

Feature-5: Using the special symbols rarely used in the legitimate URL such as “-” , “%” , ”#”, ”$” , “-”, and “^” in a domain name, close up to legitimate domain name but with special symbols to make the users believe that they deal with a legitimate site.
Rule: Any special symbol exists in the domain name, then suspicious URL else legitimate URL.
Feature-6: If number of dots in URL more than one point that will indict to a suspicious URL.
Rule: Number of dots in URL > 1 then suspicious URL else legitimate URL.
Feature-7: Check the URL if including the certificate HTTPS.
Rule: Using HTTPS then legitimate URL else suspicious URL.
Feature-8: One of Top 10 of the terms that redundant in URL phish, which extract from tests perform in and the terms are (log, pay, web, cmd, account, dispatch, free, run, net, confirm).
Rule: If URL have term from above then suspicious URL else legitimate URL.
Feature-9: If using one of top free hosting sites for phishers such as http://000webhost.com/,http://freewebhostingpro.com,http://www.freezoka.com/,,http://prohosts.org,http://phpnet.us,http://awardspace.com, http://ripway.com,http://110mb.com,http://t35.com,http://freeweb.com, http://freehostia.com.,http://superfreehost.info,
Rule: If using domain name equal any from above free hosting sites then suspicious URL else legitimate URL.

4.2.2 Google and Yahoo Response

Open URL as source code and extract the title of HTML source code, then build signature contain the URL and the web page title. The signature send as request to the Google API and Yahoo API search engines in the same time. The top 30 results received as HTML source code, then analysis source code to extract the top 30 URLs. The proposed extract URLs from HTML source code of response each search engine after that extract host names part from top 30 URL and compare the host name for the URL entered with each one of top 30 host names for Google search engine (second voter) in order to vote, if happen matching it vote as "1" but if a non - match happening its vote as"0", then check if happen matching or not with top 30 host names for Yahoo search engine (third voter).

4.3 Decision Phase

The proposed get three votes when the detect phase implemented one of test structures of a URL, the second of Google response and the third of Yahoo response. Hence, the proposed take the majority opinion (if two votes (1) the web page is legitimate otherwise the web page is phish) and display the decision to user as the message. In case the web page is phish the proposed take top site of results Google and Yahoo engines and present them to the user as an alternative legitimate site on consider the results of response the search engines come from the signature which include indicator about legitimate sites that used from a phisher to deception users. Whereas if a web page is legitimate the user implement the prevent phase which consist of the authentication steps with the server before enter the sensitive information because any detect tool may exposed to false negative error, especially with the growing threat of phishing globally so, the proposed aim to fighter Phishing by two domain the first is browsers which responsibility about access the users to sites and the second the site same especially the sensitive sites such the e-government site ,e-bank site and cloud site like this sites related with sensitive information for users, so we proposed the prevent phase, it is explained with complete algorithms in the following section.

4.4 Prevent Phase

This step proposed a method of Secret Sharing for an authentic website with generating two private keys, one for the user and another for the server, size of shares as same size the secret color image and reconstructed secret image same original image quality (see figure (2)), from known the phishing attacks implement during the login user's to his account in order to steal the sensitive information so, the proposed consist of two models: the register model and login model as following:

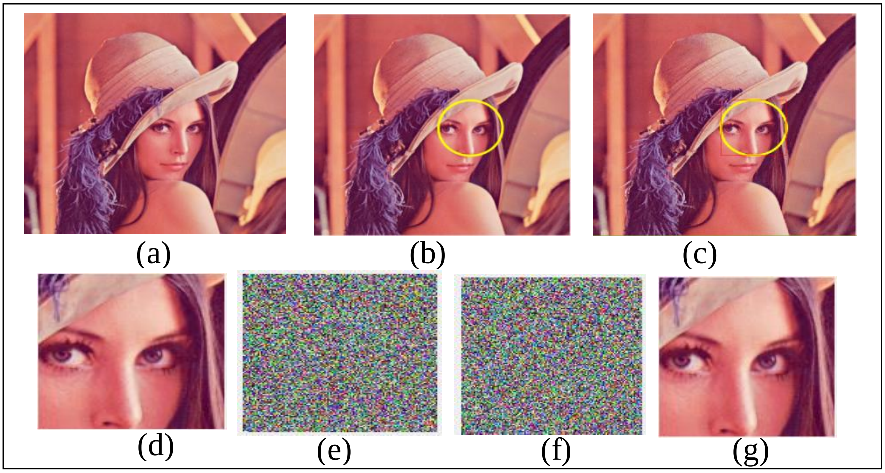

Figure 2. (a)original image(b)circle around point (x,y)(c) contour around the circle (d)the secret image(e)and(f)the two

4.4.1 Register Model

In this model the proposed assumption the user enters to site for creating a new account, the figure (3) shown the block diagram of Register proposed algorithm and steps it's as:

The Register Algorithm

Start: user registers an account on business or social media site through register username and password, then implement authenticate step between them.
Output: user has secure account.
Processing:
Step-1: User enter his email and load (the original image) any color image from his PC or web camera. **Step-2:** User performs one click within square the server displays it randomly on the loaded image to ensure the user is human.
Step-3: When a user performs one click on any point (x, y) within the image, this point will consider as the private key for the user and the center of the circle drawn, after that will determine square contour around circle.
Step-4: Cropped the contour in size (129×129) and consider it as (a secret image) between user and server.
Step-5: Generate a random key in size (129×129) with time stamp by using sum values for the point (x,y) to generate a matrix of pseudo random integers which drawn from the discrete uniform distribution on the interval [1,(x+y)].
Step-6: Convert the secret color image to the three main color images (RGB) then each image division into two share using pseudo random key and a

mathematical visual cryptography proposed in eq.(4) as alternative of standard XOR visual cryptography in order avoid expand size of shares and contrast which considers main disadvantage for standard visual cryptography.

$$\left.\begin{array}{l}\text{Share Red\_user=image\_red/random\_key + random\_key;}\\ \text{Share Red\_server=mod(image\_red,random\_key);}\\ \text{Share Green\_user=image\_green/random\_key + random\_key;}\\ \text{Share Green\_server=mod(image\_green,random\_key);}\\ \text{Share Blue\_user=image\_blue/random\_key +random\_key;}\\ \text{Share Blue\_server=mod(image\_blue,random\_key);}\end{array}\right\} \qquad (4)$$

Step-7: Enter the original image to MD5 to generate the initial value, MD5 generate 128 bits and from the first 32 bits generate four bytes (d1, d2, d3 and d4) then transform them from binary to decimal to build an initial value of x_0 by eq.5 then using the logistic map as eq.6 and eq.7 in order generate DNA rules sequence

$$x_0 = \mathrm{mod}(d_1 \oplus d_2 \oplus d_3 \oplus d_4, 256)/255 \qquad (5)$$

$$x_{n+1}=\mu x_n(1-x_n) \qquad (6)$$

Where ∈ x_n (0,1) and initial value, where µ (0,4] and in proposed we
assign value's 3.99999.

$$\text{Rule}=\text{floor}[x * 7] \qquad (7)$$

Step-8: Encode each row from share using DNA rules sequence
Step-9: Using MD5 initial value and PWLCM as eq.8 with eq.9 to generate key.

$$X_{n+1}=F_p(X_n)=\begin{cases} X_n/p, & 0< X_n < p \\ (X_n-p)/(0.5-p), & p\le X_n < 0.5 \\ F_p(1- X_n), & 0.5 \le X_n <1 \end{cases} \qquad (8)$$

Where ∈ x_n (0,1) and initial value for x generate from eq.5 and p appear as
p = 0.25678900.

$$\text{Key}=[x * 256] \qquad (9)$$

Step-10: Encode each row from key using DNA rules sequence.
Step-11: Using MD5 initial value in eq.5 and logistic map as eq.6 with eq.10 to determine DNA operation for each row in order computing the encode share from step 8 with an encoded key from Step 10 to generate the cipher share

$$\text{Operation}=\text{floor}[x*3] \qquad (10)$$

Step-12: Decode the cipher share using DNA rules sequence which generated in step 7.
Step-13: Encapsulation each share using 4-D Chen's hyperchaotic sequence which generated from the private server key (unique for each user).
Step-14: Generate two cipher color shares of six cipher shares by merging cipher user-red share with cipher user-green share and cipher user-blue share to generate the user share and same operation implemented on server shares, then the user enters his username and password. kept in server's database the user's information (email user, username, password, time stamp, server share, server

private key (initial values of hyperchaotic)) and display to the user the private user key and send (user's share, secret image) to user's email.

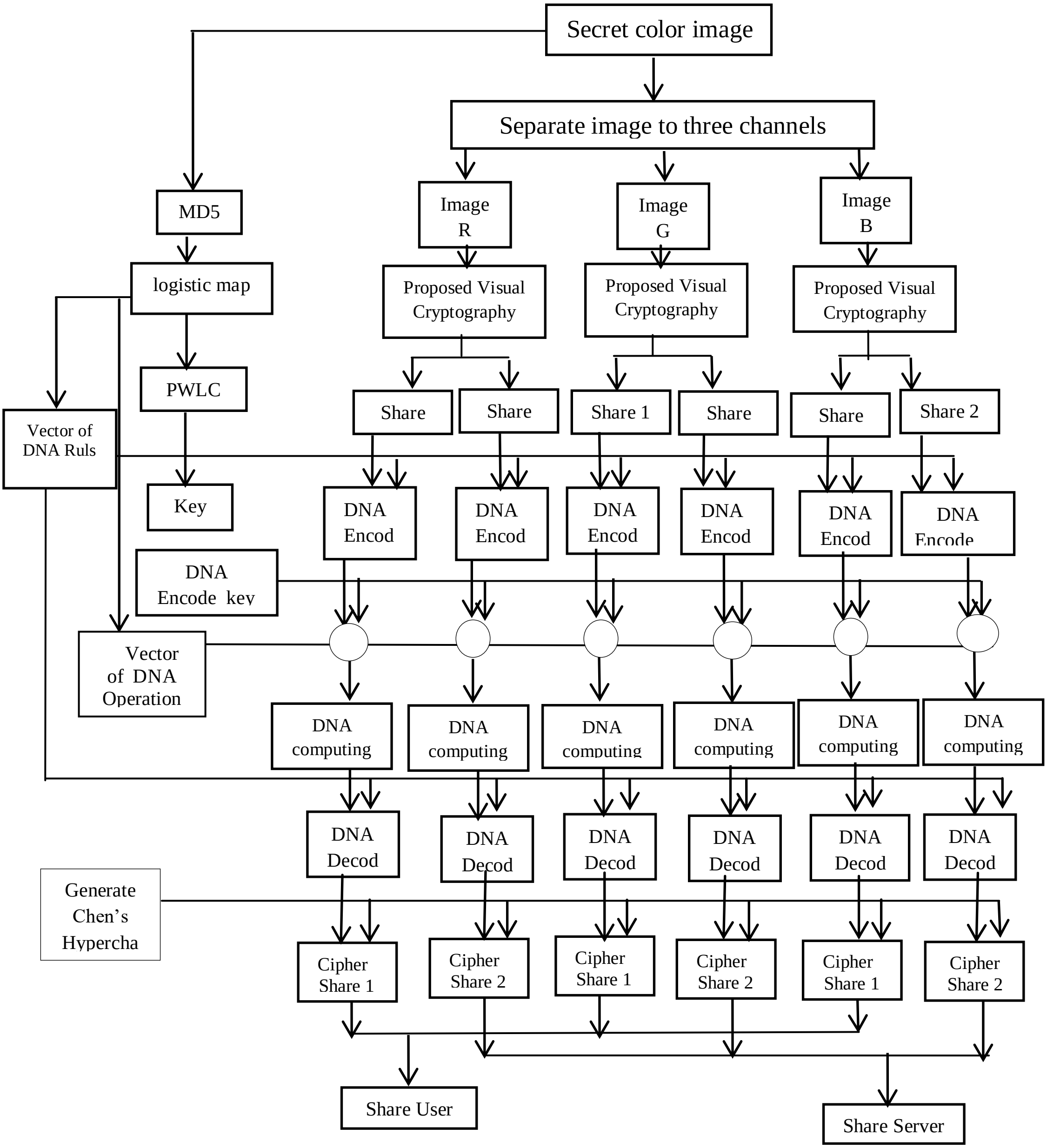

Figure 3. Block diagram of the proposed algorithm in Register Phase.

4.4.2 Login Model

Phishing attacks happening in the login phase to steal the account information for users, In the following the proposed login algorithm explanation.

The Login Algorithm

Start: User login to his account.

Output: User can login to his account after ensuring from a web page by detect operation, then user present his share and private key with loads the original image, the serve display the secret image if it's legitimate after that the user can present his sensitive information as (username, password or card cart)

Processing:

Step-1: User writes private key, then upload his share and original image.The legitimate server must retrieve his share, initial values for chen’s hyperchaotic and timestamp which dedicated to this user.

Step-2: Separate two shares to main channel color image (RGB) so six shares constructed.

Step-3: Build hyperchaotic key using initial value and decode each share.

Step-4: The original image enter to MD5 to build initial value then using logistic map (eq.6) and (eq.7) in order generate DNA rules sequence**.**

Step-5: Encode each row of share using DNA rules sequence and using MD5 initial value and PWLCM (eq.8) with (eq.9) to generate key.

Step-6: Encode each row from key using DNA rules sequence and using MD5 initial value in eq.5 and logistic map as eq.6 with eq.10 to determine DNA operation for each row in order computing the encode share from step 5 with an encoded key from Step 6 to generate the cipher share.

Step-7: Decode cipher shares using DNA rules sequence which generated in step 4.

Step-8:generate random key using the private user key and time stamp

Step-9: Using random key with six shares as the Eq.(11) then merge red and green and blue image which inverse operations for Eq.(4) to retrieve secret color image with same quality the original secret image and display it to user.

$$\left.\begin{array}{l}\text{image\_red=(Red\_user-random\_key)*random\_key+Red\_server;}\\ \text{image\_green=(Green\_user-random\_key)*random\_key+Green\_server;}\\ \text{image\_blue=(Blue\_user - random\_key)* random\_key + Blue\_server;}\end{array}\right\}\qquad (11)$$

Step-10: Visually the user ensure from the secret image which displayed if similar the secret image which chose in register phase and that has been sent to his email this make the site is legitimate and user can enter the sensitive information otherwise its phish site.

5. Results And Discussion

The Title in a web page is important because usually a page title contains the brand name of the site, the proposed extracted the web page title by use pattern matching with title tag, then build signature contain the URL and the web page title with determine two search engines to bring the top 30 results. The signature sends to search engines, then receive the top 30 results as HTML source code, then analysis source code to extract the top 30 URL. The proposed extract URLs from HTML source code after that extract host name part of top 30 URL and compare them with host name part for the URL entered. The proposed bring top 30 results and merge the title with URL in signature to reduce the false positive which happen because the newly legitimate site launch of the web have low ranking and to bring web pages which written in non-English text to the top 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** which mean the system determine legitimate link as legitimate webpage, **false positive** which mean the system determine legitimate link as phish webpage, **true negative,** which mean the system determine phish link as phish webpage and **false negative** which mean the system determine phish link as legitimate webpage. When test the proposed algorithm for 150 links of phish and 150 of legitimate, the results True Positive as *99*.4%, False Positive as *0.6*%, True Negative as 98%, and False Negative as 2%. However, many tools and methodologies have been developed to detect the Phishing attacks and to alert users orally and visually, but still the success rates of the phishing attack remain high and also the approaches related to phishing detection suffers of false negative ratio in addition to conflicting opinions in detect phishing tools about the state of URL from possible, make users *fall* as a victim of phishing for example when test the URLs which appear in our proposed as false positive and false negative using Google safe web and PhishTank the results show in table (2).

Table 2: The results test the URL

URL	State URL in data set	Our proposed	Google Web safe	PhishTank
http://envione.org	http://www.phishtank.com/phish_detail.php?phish_id=5398252 2017-12-24 T16:46:14+00:00	Legitimate Web page	No unsafe content found	Verified:Is a phish//Submission #5398252 is currently offline
http://caixafgtssaque.org	http://www.phishtank.com/phish_detail.php?phish_id=5022476 2017-05-26 T14:31:05+00:00	Legitimate Web page	No unsafe content found	Verified:Is a phish//Submission #5022476 is currently offline
https://sun-mining.com/en	http://www.phishtank.com/phish_detail.php?phish_id=5421105 2018-1-10 T07:43:23+00:00	legitimate web page	No unsafe content found	Verified: Is a phish//Submission #5421105 is currently ONLINE
https://instagram.com	One of 150 top site of alexa.com, rank 18	Phish Web page	It's hard to provide a simple safety status for sites like https://instagram.com, which have a lot of content. Sites that are safe sometimes contain some unsafe content	Voting disabled. This suspected phishing site is unavailable, probably because its host removed it Submission #3597751 is currently offline

Hence, the conflicting opinions made we proposed the prevent phase by using visual cryptography and secret sharing with private key for user and server. The proposed used the color image to authentic connection between user and server so, like this environment we perform the following tests:

5.1 Information Entropy

The information entropy to express the degree of uncertainties in the system. The information entropy can measure the distribution of gray value in the image of an idealized random image, the value of the information entropy is 8. The information entropy is defined as follows [22] :

$$H(m) = -\sum_{i=0}^{L} P(m_i)\log_2 P(m_i) \qquad (12)$$

Where m_i is the i-th gray value of L level gray image, $p(m_i)$ is the emergence probability of m_i. An effective encryption algorithm should make the information entropy tend to 8, table (3)shows information entropy for channels (R, G, B) in proposed algorithm.

Table 3: Information Entropy.

	Share User	Share Server
Red	7.988622270840382	7.988139925859467
Green	7.988483563720586	7.987543098456113
Blue	7.989578129308199	7.990025476739277

5.2 Correlation Coefficient Analysis

The horizontal, vertical, and diagonal directions are checked of all adjacent pixels from the original image and the encrypted shares. Table 4 displays the rate correlation coefficients which denote that the pixels in two color shares are uncorrelated that make the proposed has well protected of resisting statistical attack.

Table 4:The correlations for proposed algorithm.

Channels	Correlation	Share User	Share Server
Red	Diagonal	-0.008505907246551	-0.005144941891976
	Horizontal	0.006633287655889	0.017390483672535
	Vertical	0.005664386988408	-0.008189741666233
Green	Diagonal	0.002703783633764	0.004126214321466
	Horizontal	-0.010549143540311	-0.014473871054005

	Vertical	-0.014922927414818	-0.004113191383055
Blue	Diagonal	-0.009616273420571	2.724765763235e-04
	Horizontal	-0.011977383517769	0.013695075191974
	Vertical	-0.005458114846746	-0.003581762275749

5.3 NPCR and UACI

The attacker may be able to detect a meaningful relationship between the plain image and the cipher image by making a small change of the encrypted image then watch the change of the result. If one pixel change in the plain image can cause a considerable change in the cipher image, in this way the differential attack would become very ineffective and practically useless. The number of pixels change rate (NPCR) and unified average changing intensity (UACI) two common measures to test the influence of one pixel change on the whole encrypted image by encrypting two images with one-pixel value difference in their original images are recorded as C1(i, j) and C2(i, j) were NPCR to measure the percentage of different pixel numbers between two images as eq.(13), while the eq.(11) refer to UACI to measure the average intensity of the differences between two images [23]. Table (5) shown NPCR and UACI for proposed algorithm and other methods.

$$NPCR = \frac{\sum_{ij} D(i,j)}{Width \times Hight} \times 100\% \qquad (13)$$

$$NPCR = \frac{\sum_{ij} D(i,j)}{Width \times Hight} \times 100\% \qquad (14)$$

Table 5: NPCR and UACI for proposed algorithm and other approach.

		Red	Green	Blue
Proposed algorithm	NPCR(%)	99.7175	99.6334	99.68150
	UACI(%)	33.8095	33.5372	33.66532
Wu's algorithm[23]	NPCR(%)	99.6101	99.6136	99.6141

	UACI(%)	33.4695	33.4643	33.4665
Wang's algorithm [24]	NPCR(%)	99.6006	99.6178	99.5975
	UACI(%)	33.4418	33.5298	33.4927
Murillo-Escobar's algorithm [25]	NPCR(%)	99.63	99.60	99.61
	UACI(%)	33.31	33.34	33.43
Liu's algorithm[26]	NPCR(%)	99.6231	99.6338	99.6170
	UACI(%)	33.4747	33.5683	33.3382

5.4 Histogram Analysis

Figure (4) shows the histogram result when test lena image under the proposed algorithm.

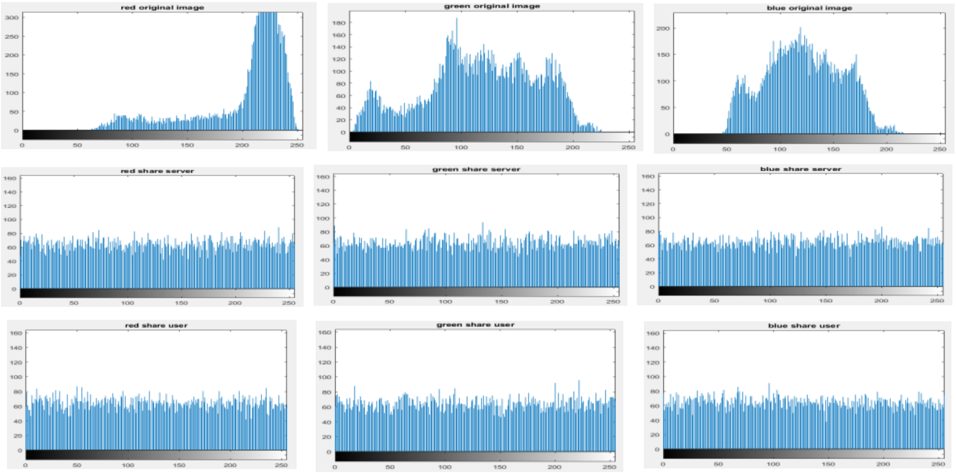

Figure 4. Shows histograms of original secret image {Red, Green and Blue} in first row and Red, Green and Blue histograms images for encrypted color shares in second and third rows.

5.5 Hyper-Chaotic Key

In proposed algorithm, the secret key for server representation of the initial values of the Chen's hyper-chaotic system. Thus, t1, t2, t3, and t4 used as server key and supposed the t1 use value equal to 0.300000000000001 instead of 0.3 in order decode the encrypted shares that leads to generate different shares and because each variable from four initial values have sensitivity to change its value reach to 10^{15} , the proposed algorithm employ this sensitivity with exploit new initial values for each new registration. In other word in each new registration request the server choose one variable randomly [t1, t2, t3, and t4] to change its initial value based on employing the random number generator based on the current time with range 10^{13} to build new initial values for the Chen's hyper-chaotic system.

6 Conclusion

This paper proposed effective algorithm to detect and prevent the phishing attacks based on not only analysis the contains of URL by check specific features if found or not and build the signature from merging the URL with web page title which extracted from HTML source to relative information from Google and Yahoo search engines but also on proposed the prevent phase based on new visual cryptography encapsulated with secret share. The proposed able to detect the web page newly launched over the web and non-English text web page depended on majority vote with give results compared with Google web safe and PhishTank performance. This paper proposed scheme to deal with phishing threat by detect operation to determine the site and prevent operation to authentic the connection with customers's to avoid economic losses of the two parties.

References

[1] Mohammad, R., Fadi, A., & McCluskey, L. (2014). “Predicting phishing websites based on self-structuring neural network”. Neural Computing and Applications, 25(2), pp(443-458).

[2] Mohammad, R., Fadi, A., & McCluskey, L. (2015, February). “Tutorial and critical analysis of phishing websites methods”. Computer Science Review, 17, pp(1-24).

[3] APWG (Anti-Phishing Working Group). 2017. phishing trends reports.

[4] Buber, E., Demir, Ö., & Sahingoz, O. (2017, September). “Feature selections for the machine learning based detection of phishing websites”. In Artificial Intelligence and Data Processing Symposium (IDAP), 2017 International (pp. 1-5), IEEE.

[5] Dunlop, M., Groat, S., & Shelly, D. (2010, May). “Goldphish: Using images for content-based phishing analysis”. In Internet Monitoring and Protection (ICIMP), 2010 Fifth International Conference on (pp. 123-128), IEEE.

[6] Ashwaq, T., Hashim, & George, L. E. (2013). “Secret Image Sharing Based on Wavelet Transform”. In International Conference on Information Technology in Signal and Image Processing, Mumbai, India (pp. 324-332).

[7] Ram, B., Andrew, H., & Quingzhong, L. (2014, Jun). “Learning to detect phishing URLs”. International Journal of Research in Engineering and Technology, 3(6), pp(11-24).

[8] Al-Khalid, R. I., Al-Dallah, R. A., Al-Anani, A. M., Barham, R. M., & Hajir, S. I. (2017, January). “A Secure Visual Cryptography Scheme Using Private Key with Invariant Share Sizes”. Journal of Software Engineering and Applications, 10(01), pp (1-10).

[9] Varshney, G., Misra, M., & Atrey, P. K. (2016, December). “Improving the accuracy of Search Engine based anti-phishing solutions using lightweight features”. In Internet Technology and Secured Transactions (ICITST), 2016 11th International Conference for (pp. 365-370). IEEE.

[10] Amiri, I. S., Akanbi, O. A., & Fazeldehkordi, E. (2014,). “A Machine-learning Approach to Phishing Detection and Defense”. Syngress.

[11] Sathishkumar, G. A., & Sriraam, D. N. (2011, March). “Image encryption based on diffusion and multiple chaotic maps”. International Journal of Network Security & Its Applications (IJNSA), 3(2), pp(181-194).

[12] Guodong, Y., & Xiaoling, H. (2017, April). “An efficient symmetric image encryption algorithm based on an intertwining logistic map”. Neurocomputing, 251, pp(45-53).

[13] Hua, W., & Liao, X. (2017). “A secret image sharing scheme based on piecewise linear chaotic map and Chinese remainder theorem”. Multimedia Tools and Applications, 76(5), pp(7087-7103).

[14] Chen, A., Lu, J., Lü, J.,& Yu, S. (2006). “Generating hyperchaotic Lü attractor via state feedback contro”l. Physica A: Stat. Mech. Appl. 364(C), pp(103–110).

[15] Zheng, W., Wang, F. Y., & Wang, K. (2017, October). “An ACP-based Approach to Color Image Encryption Using DNA Sequence Operation and Hyper-chaotic System”. 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC).pp(461- 466).

[16] Azad, S., & Pathan, A. (2014). “Practical Cryptography: Algorithms and Implementations Using C++”. CRC Press.

[17] Enayatifar, R., Abdullah, A. H., & Isnin, I. F. (2014). “Chaos-based image encryption using a hybrid genetic algorithm and a DNA sequence”. Optics and Lasers in Engineering, 56, pp(83-93).

[18] Niyat, A. Y., Hei, R. M. H., & Jahan, M. V. (2015, October). “A RGB image encryption algorithm based on DNA sequence operation and hyper-chaotic system”.Conference Paper.

[19] Yan, X., Lu, Y., Chen, Y., Lu, C., Zhu, B., & Liao, Q. (2017, May). “Secret image sharing based on error-correcting codes”. In Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), 2017 IEEE 3rd International Conference on (pp. 86-89). IEEE.

[20] Liu, F., & Yan, W. Q. (2014). “Visual Cryptography for Image Processing and Security” (Vol. 2). New York: Springer.

[21] Aburrous, M., Hossain, M. A., Dahal, K., & Thabtah, F. (2010). “Intelligent phishing detection system for e-banking using fuzzy data mining”. Expert systems with applications, 37(12), pp(7913-7921).

[22] Mehdi, S. A., & Kareem, R. S. (2017). “Using Fourth-Order Runge-Kutta Method to Solve Lü Chaotic System”. American Journal of Engineering Research (AJER), 6 (1), pp(72-77).

[23] Wu, X., Wang, D., Kurths, J., & Kan, H. (2016). “A novel lossless color image encryption scheme using 2D DWT and 6D hyperchaotic system”. Information Sciences, 349, pp(137-153).

[24] Yao, W., Wu, F., Zhang, X., Zheng, Z., Wang, Z., Wang, W., & Qiu, W. (2016, November). “A Fast Color Image Encryption Algorithm Using 4-Pixel Feistel Structure”. PloS one, 11(11), pp(1-30).

[25] Murillo-Escobar, M. A., Cruz-Hernández, C., Abundiz-Pérez, F., López-Gutiérrez, R. M., & Del Campo, O. A. (2015). “A RGB image encryption algorithm based on total plain image characteristics and chaos”. Signal Processing, 109, pp(119-131).

[26] Liu, H., & Kadir, A. (2015). “Asymmetric color image encryption scheme using 2D discrete-time map”. signal processing, 113, pp(104-112).

كشف ومنع التصيد بالاستناد على نهج هجين

أ.م.د.هالة بهجت عبدالوهاب*

م.م.ذكرى محمد عبد*

المستخلص: التصيد هو أحد التهديدات على شبكة الإنترنت، ويعتبر هجوم التصيد ناجحا إذا كان المستخدم بدون دراية يقدم معلوماته الشخصية للمهاجم هذه الهجمات اصبحت شعبية في الآونة الأخيرة. وقد تواجه البلدان النامية مثل العراق تهديدات على الإنترنت مثل التصيد الاحتيالي. تهدف هذه الورقة إلى اقتراح نظام قائم على خوارزمية مقترحة لكشف ومنع هجمات التصيد الاحتيالي, هذا المقترح يحلل البنية الهيكلية وشفرة المصدر للرابط ويعتمد على محركات البحث جوجل وياهو في عملية الكشف. اما عملية المنع فتعتمد على طريقة مقترحة من التقاسم السري لإنتاج اثنين من الاسهم واحد للمستخدم والاخر للخادم مع مفاتيح خاصة لهما. النتائج التجريبية التي نفذت على 300 رابط أظهرت معدل الإيجابية الصحيحة 99.4٪ ومعدل الإيجابية الكاذبة 0.6٪ مع السلبية الصحيحة 98٪ والسلبية الكاذبة 2٪ ثم تم مقارنة النتائج الكاذبة مع التقارير التي صدرت من Googl Web Safe and Phish Tank اضافة الى بناء اسهم سرية غير موسعة قادرة على اعادة تشكيل نفس جودة الصورة الاصلية .

الكلمات المفتاحية: كشف التصيد, التصيد الاحتيالي , التقاسم السري, توسيع المتصفح.

*الجامعة التكنولوجية -قسم علوم الحاسوب