Phishing Detection Using Visual Cryptography

Deepshika D. J.,1 Murugesan M2
1UG Scholar, Dhanalakshmi College of Engineering, Tamilnadu, India
2Associate Professor, Dhanalakshmi College of Engineering, Tamilnadu, India

ABSTRACT

Phishing is an attempt by an individual or a group to thief personal confidential information such as passwords, credit card information etc from unsuspecting sufferer for burglary, financial gain and other criminal activities. The first defense should be strengthening the authentication mechanism in a web application. A simple username and password based authentication is not sufficient for web sites providing critical financial transactions. Here we have advised a new way for phishing websites classification to solve the problem of phishing. Phishing websites involves a collection of key within its content-parts as well as the browser-based security indicators provided along with the website. The use of images is try to keep the privacy of image captcha by dissolve the original image captcha into two shares that are stored in separate database servers such that the original image captcha can be acknowledged only when both are available together the individual sheet images do not confess the status of the original image captcha. Once the original image captcha is announced to the user it can be used as the password. Several solutions have been suggested to handle phishing.

Keywords: Phishing Websites, Visual Cryptography, Image Processing, Antiphishing.

I. INTRODUCTION

Online transactions are nowadays become very common and there are various attacks present behind this. In these types of various attacks, phishing is determined as a major security threat and new inventive ideas are rising with this in each second so defending mechanism should also be so powerful. Thus the security in these cases be very high and should not be easily tractable with implementation easiness. Today, most applications are only as secure as their basic system. Since the design and technology of middleware has improved regularly, their detection is a difficult problem. As a result, it is nearly impossible to be sure whether a computer that is connected to the internet can be considered believable and secure or not. Phishing blackmail are also becoming a problem for online banking and e-commerce users. The query is how to handle applications that needs a high level of surveillance. Phishing is a form of online identity theft that plans to take responsive information such as online banking passwords and credit card information from users. Phishing scams have been receiving huge press coverage because such attacks have been expanded in number and elegance. one definition of phishing is given as “it is a criminal activity using social engineering techniques. Phishers attempt to dishonestly acquire precise information, such as passwords and credit card details, by pretend as a trustworthy person or business in an electronic communication”. The conduct of identity theft with this acquired sensitive information has also become easier with the use of technology and identity theft.
can be described as “a crime in which the impostor obtains key pieces of information such as Social Security and driver’s license numbers and uses them for his or her own gain”. Phishing attacks built upon a mix of technical duplicity and social engineering practices. In the majority of cases the phisher must get the person to willingly perform a series of actions that will provide access to secret information. Communication channels such as email, WebPages, IRC and instant messaging services are popular. In all cases the phisher must act like a trusted source for the user to believe. To date, the most successful phishing attacks have been initiated by email – where the phisher impersonates the sending authority So here introduces a new method which can be used as a safe way against phishing which is named as "A novel approach against Anti-phishing using visual cryptography". As the name suggests ,here the website cross checks its own identity and confirms that it is a trusted website (to use bank transaction, E-commerce and online booking system etc.) before the end users and make the both the sides of the system secure as well as an authenticated one. The approach of image processing and an improved visual cryptography is used. Image processing is a method of converting an input image and to get the output as either enhanced form of the same image and/or quality of the input image. Visual Cryptography (VC) is a approach of encrypting a secret image to shares, such that assembling a acceptable number of shares that confess the secret image.

II. LITERATURE SURVEY

THE FOG COMPUTING PARADIGM: SCENARIOS AND SECURITY ISSUES
AUTHORS: I. Stojmenovic and S. Wen
YEAR: 2014
DESCRIPTION

Fog Computing is a paradigm that extends Cloud computing and services to the edge of the network. Identical to Cloud, Fog provides data, compute, storage, and application services to end-users. Elaborate the motivation and advantages of Fog computing, and analyze its applications in a series of real scenarios, such as Smart Grid, smart traffic lights in vehicular networks and software defined networks. Security and privacy issues are further admit according to current Fog computing standard. As an example, study a typical attack, man-in-the-middle attack, for the discussion of security in Fog computing. Consider the private features of this attack by reviewing its CPU and memory utilization on Fog device.

MERITS
Fog computing advantages for services in several domains, and provide the analysis of the state-of-the art and security issues in current paradigm

DEMERTS
Portability between Fog nodes, and between Fog and Cloud, can be checked.

ANTIPHISHING THROUGH PHISHING TARGET DISCOVERY
AUTHORS: L. Wenyin, G. Liu, B. Qiu, and X. Quan
YEAR: 2012
DESCRIPTION
Phishing attacks are growing in both volume and sophistication. The ant phishing method described here collects WebPages with either a direct or indirect association with a given suspicious webpage. This enables the discovery of a webpage’s so-called “parasitic” community and then ultimately its phishing target that is, the page with the strongest parasitic relationship to the suspicious webpage. Discovering this goal lets users conclude whether the given webpage is a phishing page.
MERITS
If the advertising link is detected as phishing, and its target is found, the target owner can be notified and take necessary action.

DEMERITS
When the attack begins, the time these solutions require to verify the attack demonstrates that they don’t effectively protect users.

A PAGERANK BASED DETECTION TECHNIQUE FOR PHISHING WEB SITES
AUTHORS: A. N. V. Sunil and A. Sardana
YEAR: 2012
DESCRIPTION
Phishing is an attempt to acquire one’s information without user’s knowledge by tricking him by making similar kind of website or sending emails to user which looks like legitimate site or email. Phishing is a cyber attack, which is causing severe loss of savings to the user, due to phishing attacks online transaction users are rejecting. To design and implement a new technique to detect phishing web sites using Google's Pagerank. Google provides a Pagerank value to every website in the web. The Pagerank value and other features are used here to segregate phishing sites from normal sites. The collected a dataset of 100 phishing sites and 100 legitimate sites for our use. By using Google Pagerank technique 98% of the sites are positively detected, showing only 0.02 false positive rate and 0.02 false negative rate.

MERITS
This technique will easily classify the phished URL’s. Phishing sites will have very less GTR value so they can be easily identified as phished sites by using the values of this heuristic and other five heuristics.

DEMERITS
The system can be implemented by adding more heuristics to the technique proposed to attain high accuracy rate to classify the phishing sites from the legitimate sites.

SOFTWARE-DEFINED NETWORK FUNCTION VIRTUALIZATION: A SURVEY
AUTHOR: Y. Li and M. Chen
YEAR: 2015
DESCRIPTION
Diverse proprietary network appliances increase both the capital and operational expense of service providers, meanwhile causing problems of network ossification. Network function virtualization (NFV) is recommended to address these issues by carrying out network functions as pure software on commodity and general hardware. NFV allows adjustable facilities, arrangement, and centralized management of virtual network functions. Along with SDN, the software-defined NFV architecture further offers lively traffic steering and joint increase of network functions and resources. This architecture gains a wide range of applications (e.g., service chaining) and is becoming the superior form of NFV. In this survey, a thorough investigation of the development of NFV under the software-defined NFV architecture, with an emphasis on service chaining as its application. First recommend the software-defined NFV architecture as the condition of the art of NFV and present connection between NFV and SDN. Then, provide a historic view of the involvement from middle box to NFV. Finally, Introduce significant challenges and relevant solutions of NFV, and discuss its future research directions by different application domains.

MERITS
Introduce NFV its relationship with SDN. Also look at the history of NFV, presenting how middle boxes evolve to virtual network functions.

DEMERITS:
The obtained higher resource utilization will introduce less investigation on the hardware equipments.
PHISHING WEB PAGE DETECTION

AUTHOR: L. Wenyin, G. Huang, L. Xiaoyue, X. Deng, and Z. Min

YEAR: 2005

DESCRIPTION
An approach to detection of phishing WebPages based on visual similarity is proposed, which can be utilized as a part of an enterprise solution to antiphishing. A legitimate webpage owner can use this approach to search the Web for suspicious WebPages which are visually similar to the true webpage. The approach first decomposes the WebPages into salient (visually distinguishable) block regions. The visual similarity between two WebPages is then evaluated in three metrics: block level similarity, layout similarity, and overall style similarity. A webpage is stated as a phishing suspected if any of them (with regards to the true one) is above than its comparable preset opening. Preliminary experiments show that the approach can successfully detect those phishing WebPages with few false alarms at a speed adequate for online application.

MERITS:
A true webpage owner can use this approach to detect phishing WebPages.

DEMERITS:
This approach is limited to detection of this kind of phishing attacks.

ANFIS: ADAPTIVE-NETWORK-BASED FUZZY INFERENCE SYSTEM

AUTHOR: J.-S. R. Jang

YEAR: 1993

DESCRIPTION
The architecture and learning procedure underlying ANFIS (adaptive-network based fuzzy inference system) is presented, which is a fuzzy inference system implemented in the framework of adaptive networks. By using a hybrid learning procedure, the proposed ANFIS can construct an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and stipulated input-output data pairs. In the imitation, the ANFIS architecture is engaged to model nonlinear functions, recognize nonlinear components on-linearly in a control system, and conclude a confused time series, all yielding remarkable results. Comparisons with artificial neural networks and earlier work on fuzzy modelling are listed and discussed. Other development of the suggested ANFIS and assuring applications to automatic control and signal processing are also recommended.

MERITS:
By employing a hybrid learning procedure, the proposed architecture can refine fuzzy if-then rules obtained from human experts to describe the input-output behavior of a complex system.

DEMERITS:
Important issue in the training of ANFIS is how to preserve the human-plausible features such as bell-shaped membership functions.

TEXTUAL AND VISUAL CONTENT-BASED ANTI-PHISHING: A BAYESIAN APPROACH


YEAR: 2011

DESCRIPTION
A novel framework using a Bayesian approach for content-based phishing web page detection is presented. Our model takes into report of textual and visual contents to analyse the likeliness between the secured web page and malicious web pages. A text classifier, an image classifier, and an algorithm merging the results from classifiers are brought in. An outstanding feature of this exploration of a Bayesian model to estimate the matching threshold. This is required in the classifier for deciding the class of the web page and determining whether the web page is phishing or not. In the text classifier, the naive Bayes rule is used to find out the probability
that a web page is phishing. In the image classifier, the earth mover's distance is employed to measure the visual similarity, and our Bayesian model is designed to determine the threshold. In the data fusion algorithm, the Bayes theory is used to synthesize the classification results from textual and visual content. The influence of our suggested method was analysed in a large-scale dataset gathered from real phishing cases. Experimental results demonstrated that the text classifier and the image classifier the designed deliver promising results, the fusion algorithm outperforms either of the individual classifiers, and our model can be adapted to different phishing cases.

**MERITS**
The new features of this framework can be represented by a text classifier, an image classifier, and a fusion algorithm. Based on the textual content, the text classifier is able to classify a given web page into corresponding categories as phishing or normal.

**DEMERITS**
It is not worth for content-based model can be easily embedded into current industrial anti-phishing systems.

**NEURO-FUZZY MODELING AND CONTROL**
**AUTHOR:** J.-S. Jang and C.-T. Sun
**YEAR:** 1995
**DESCRIPTION**
Fundamental and advanced developments in neuro-fuzzy synergisms for modelling and control are reviewed. The important part of neuro-fuzzy synergisms is derived from a framework called adaptive networks, which consolidates both neural networks and fuzzy models. The fuzzy models that comes under the framework of adaptive networks are called Adaptive-Network-based Fuzzy Inference System (ANFIS), which contains some merits over neural networks. The design methods for ANFIS in both modelling and control applications. Current issues and future enhancements for neuro-fuzzy approaches are also given.

**MERITS**
Rich literature of optimization, which offers many better gradient-based optimization routines, such as quadratic programming and conjugate gradient descent.

**DEMERITS**
Need to search for better learning algorithms hold equally true for both neural networks and fuzzy models.

**AN EMPIRICAL ANALYSIS OF PHISHING BLACKLISTS**
**AUTHOR:** S. Sheng, B. Wardman, G. Warner, L. F. Cranor, J. Hong, and C. Zhang
**YEAR:** 2009
**DESCRIPTION**
The effectiveness of phishing black lists. Used 191 fresh phish that were less than 30 minutes old to conduct two tests on eight anti-phishing toolbars. Found that 63% of the phishing attacks in our dataset existed less than two hours. Blacklists were useless when defending users at first, as most of them caught less than 20% of phish at hour zero. Also found that blacklists were renewed at varied speeds, and in varied coverage, as 47% - 83% of phish were appeared on blacklists 12 hours from the initial test. Found that two tools using heuristics to complement blacklists caught significantly more phish initially than those using only blacklists. Still, it took a long time for phish identified by examiners to appear on blacklists. Finally, tested the toolbars on a set of 13,458 legitimate URLs for false positives, and did not find any instance of mislabelling for either blacklists or heuristics. These verdict and conferring ways anti-phishing tools can be enhanced.
The window of opportunity for defenders can be defined as the length of the phishing campaign plus the time lapse between the time a user receives a phishing email and the time the user opens the email.

**MERITS**
URLs came from a single anti-spam vendor; therefore the URLs received may not be representative of all phishing.

**DEMERITS**

**FIGHTING AGAINST PHISHING ATTACKS: STATE OF THE ART AND FUTURE CHALLENGES**
**AUTHOR:** B. B. Gupta, A. Tewari, A. K. Jain, and D. P. Agrawal
**YEAR:** 2016

**DESCRIPTION**
In the last few years, phishing scams have rapidly grown posing huge threat to global Internet security. Today, phishing attack is one of the most common and serious threats over Internet where cyber attackers try to steal user’s personal or financial credentials by using either malwares or social engineering. Identification of phishing attacks with high efficiency has always been an issue of great interest. Latest improvements in phishing detection techniques have led to various new techniques, mainly planned for phishing detection where certainty is extremely important. Phishing problem is broadly prevailing as there are many ways to carry out such an attack, which indicate that one solution is not sufficient to explain it. Two main issues are addressed. First, discuss in detail phishing attacks, history of phishing attacks and motivation of attacker behind performing this attack. In addition, also provide taxonomy of various types of phishing attacks. Second, provide taxonomy of various solutions proposed in the literature to detect and defend from phishing attacks. In addition, also discuss various issues and challenges faced in dealing with phishing attacks and spear phishing and how phishing is now targeting the emerging domain of IoT. Consider variety of tools and datasets that are used by the scholars for the assessment of their approaches. This gives us better understanding of the problem, the present outcome space and future research scope to skilfully deal with such attacks.

**MERITS**
The machine learning techniques give the best results as compared to other techniques as they are able to mitigate zero-hour phishing attacks better than the other.

**DEMERITS**
Both sender and receiver must use the same technique. Cannot be secured from MITM attacks. High FP (False positive) rates and bandwidth requirements.

**III. SYSTEM ANALYSIS**

**EXISTING SYSTEM**
Phishing web pages are fake web pages that are created by venomous people to imitates Web pages of real web sites. Most of these kinds of web pages have high visual likeliness to fraud their victims. Some of these kinds of web pages look accurately like the real ones. Users of phishing web pages may reveal their bank account, password, credit card number, or other important information to the phishing web page owners. It involve techniques such as cheat customers through email and spam messages, man in the middle attacks, installation of key loggers and screen captures.

**DISADVANTAGES**
These popular technologies have several drawbacks: Blacklist-based technique with low false alarm probability, but it cannot detect the websites that are not in the blacklist database. Because the life cycle of phishing websites is short and the creation of blacklist has a more lag time, the precision of blacklist is not too high.
Heuristic-based anti-phishing method, with a increased probability of false alarm, and it is easy for the phisher to use technical means to avoid the heuristic characteristics detection.

Similarity assessment based technique is time-consuming. It takes too long to analyze a pair of pages, so using the method to identify phishing websites on the client side is not applicable. And there is low precision rate for this technique that depends on many factors, such as the text, images, and similarity measurement.

**PROPOSED SYSTEM**

The technique of image processing and an improved visual cryptography is used. Image processing is a method of dealing with an input image and to get the output as either enhanced form of the same image and/or tendency of the input image. In Visual Cryptography (VC) an image is divided into shares and in order to disclose the original image correct number of shares should be connected.

VCS is a cryptographic approach that permits the encryption of visual information such that decryption can be achieved using the human visual system. We can attain this by one of the following access schemes

- **(2, 2)**- Threshold VCS scheme- This is a easiest threshold action that takes a secret key message and encrypts it in two different parts that shows the secret image when they are combined.

- **(n, n)** -Threshold VCS scheme-This approach encrypts the secret image to n shares such that when all n of the shares are overlaid will the secret image be shown.

- **(k, n)** Threshold VCS scheme- This scheme encrypts the secret image to n shares such that when any group of at least k shares are combined the secret image will be revealed.

In the case of (2, 2) VCS, each pixel P in the actual image is encrypted into two sub pixels called shares. They are shares of a white pixel and a black pixel. Note that the number of shares for a white and black pixel is randomly determined (there are two choices available for each pixel). Neither share gives any clue about the original pixel since different pixels in the secret image will be encrypted using random ways. When the two shares are combined, the value of the original pixel P can be revealed. If P is a black pixel, we get two black sub pixels; if it is a white pixel, we get one black sub pixel and one white sub pixel.

**ADVANTAGES**

For phishing detection and prevention, we are suggesting a new method to identify the phishing website. Our technique is based on the Anti-Phishing Image Captcha validation scheme using visual cryptography.

It safeguards password and other confidential information from the phishing websites.

URL on the address bar of your webpage begins with "https"; the alphabet at the end of "https" which means 'secured'.

Look for the padlock symbol in the address bar or the status bar (mostly in the address bar) but not inside the web page display area. Check the authentication by selecting the padlock.

**IV. RESULTS AND DISCUSSION**

**4.1 REQUIREMENTS SPECIFICATION**

**4.1.1 SOFTWARE REQUIREMENTS**

Operating System : Windows XP or Higher  
Languages used : Java (JSP, Servlet), HTML  
Tools : JDK 1.7, Net Beans 7.0.1, SQLyog  
Backend : My SQL

**4.1.2 HARDWARE REQUIREMENTS**

Processor : Pentium Dual Core 2.3 GHz  
Hard Disk : 250 GB or Higher  
Ram : 1 GB (Min)
ALGORITHM:
RANDOM NUMBER:
Java builds a number of functions that you can use in your programs. However, Java runs with a less amount of methods already available. This helps with running time and reduces extra lines of code. When you want to use a certain feature, such as getting input from the user or creating a random number, you need to import that preprocessor into your code. We use the statement to do this. The tool we’ll be using is the Random class, which is part of Java’s utility (util) library.

We can create random number using rand object. We should provide some information to rand object so that the number generated varies. If we do not provide the information then the compiler always generates the same output.

Import java.util.Random;

We will need to set our constraints; in this case we need a number between 1 and 100. Create two integer variables for this, like this:
Int max=100;
Int min=0;

Next we can create the random number generator. Random is a class in Java, with its own functions. We can create an instance of this class and use all of these features. In order to create a new instance of Random, this code is used.

At last, we can create our random number. The Random class shows a method called nextInt(int n), which creates a random number between 0 and the number specified (n). We need only numbers between 1 and 100. It may look a little puzzling, but we’ll go through it. This code is to create a random number between 1 and 100 and save it to a new integer, showMe:

```java
int showMe = min + randomNum.nextInt(max);
```

Every time we run the program a different number is displayed between 1 and 100. Here is the final, complete code:

```java
public static void main(String[] args) {
    //what is our range?
    int max = 100;
    int min = 1;

    //create instance of Random class
    Random randomNum = new Random();

    int showMe = min + randomNum.nextInt(max);
    System.out.println(showMe);
}
```

V. SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE

SYSTEM IMPLEMENTATION

MODULES

Registration With Secrete Code
Image captcha Generation
Shares Creation (VCS)

Login Phase

Module Description
Registration With Secret Code:

In this registration phase, the user details username, password, email-id, address, and a key string (password) is attained from the user at the time of registration for the protected website. The key string can be a mixture of alphabets and numbers to provide more protected environment. This string is merged with randomly created string in the server.

Image captcha Generation:

A key string is changed into image using java classes BufferedImage and Graphics2D. The image dimension is 260 multiple of 60. Text color is red and the background color is white. Text font is about by Font category in java. After image creation it will be written into the user key folder in the server by using ImageIO class.

Shares Creation (VCS):

The image captcha is divided into two shares that one of the share is given to the user and the other share is kept in the server. The user's share and the original image captcha sent to the user for further authentication during login phase. The image captcha is also kept in the actual database of any authenticated website as secured data.

Login Phase:

When the user logs in by entering his private information for using his account, then first the user is asked to give his username (user id). Then the user is requested to enter his share that is unbroken with him. This share is shipped to the server wherever the user's share and share that is keep within the info of the web site for every user, is stacked together to produce the image captcha. The image captcha is displayed to the user. Here the top user will check whether or not the displayed image captcha matches with the captcha created at the time of registration. The end user is needed to enter the text displayed within the image captcha and this may serve the aim of arc a num and victimization this, the user can log in into the website. Using the username and image captcha generated by stacking 2 shares one will verify whether or not web site the web site is genuine/secure web site or a phishing website.

Product Perspective

This product is combination of our main components, namely Image processing and visual cryptography, the web portal, web services and the JEE application. The main objective is predicting the phishing sites supported visual cryptography.

VI. CONCLUSION

Currently phishing attacks square measure thus common as a result of it will attack globally and capture and store the users’ steer. This information is used by the attackers which are indirectly involved in the phishing process. Phishing websites similarly as human users may be simply known mistreatment our planned “Anti-phishing framework supported Visual Cryptography”. The proposed methodology preserves confidential information of users. Verifies whether or not web site the web site could be a genuine/secure web site or a phishing website. If web site the web site could be a phishing web site (website that’s a pretend one simply the same as secure website however not the secure website), then therein scenario, the phishing website can’t display the image captcha for that specific user (who wants to log in into the website) due to the fact that the image captcha is generated by the stacking of 2 shares, one with the user and the other with the
actual database of the website. The planned methodology is additionally helpful to stop the attacks of phishing internet sites on money web portal, banking portal, on-line looking market.

VII. FUTURE ENHANCEMENT

In future we are able to increase the safety by adding several algorithms to encode the image. Encryption Phase contains many algorithms like Blowfish, Splitting and Rotating algorithm and (2,2) Visual Cryptography Scheme. First the “Blowfish Algorithm” is applied to the initial image captcha then the image captcha is split into several blocks and rearranged. After the image captcha blocks are rearranged, the “Splitting and Rotating Algorithm” is applied to the image captcha, and then the rearranged blocks are rotated. Then the rearranged and rotated blocks are combined. Then (2, 2) VCS theme is applied to the combined blocks. This theme is employed to divide the encrypted image captcha into 2 shares supported white and black pixels. When the 2 sub components are identical blocks it considers as a white pixel. Likewise once the 2 sub components are completely different the initial component is taken into account as black pixel. This VCS theme adds a lot of complication to the image captcha.

VIII. REFERENCES


Journal URL : http://ijsrst.com/IJSRST196248