# Preventing Phishing Attack Using Visual Cryptography

Prof. Prajwal Gaikwad[*1], Akanksha Chimote[#2], Shubhangi Bhaladhare[#3], Mayuri Dahiwal[#4], Ankita Ambade[#5]

[*1]*Assistant Professor at Computer Department*

[1]prajwalg20@gmail.com

[#]*Computer Engineering, Pune University*
*AISSM'S Institute Of Information Technology, Pune, Maharashtra, India*

[2]akankshachimote@gmail.com

[3]shubhangibhaladhare92@gmail.com

[4]mayuridahiwal@gmail.com

[5]ankitaambade@gmail.com

*Abstract*— **Phishing is a method that used by scammer to collect personal confidential information such as password, credit card information from unsuspecting user using forged web pages. So for detection and prevention from phishing we have proposed a new approach named "Anti-phishing based on Visual Cryptography". Unlike traditional cryptography, visual cryptography uses simple algorithm. This technique allows the visual information to be encrypted without any complex cryptographic algorithm so that decryption can be performed by human only. In this scheme an image is divided in two shares where one secret CAPTCHA image resides with user and other reside in server. During authentication process a genuine server forward his share. Both share stack together to reconstruct image and it must be same and prone to character recognition based attack. The original image CAPTCHA known to the user it can be used for further process.**

*Keywords*— **Phishing, Visual Cryptography, Image CAPTCHA, Shares, Security**

## I. Introduction

Communication is the basic need of human being in today's world. The fastest way to communicate all over world is internet. One of the important use of internet is online banking, E-commerce, financial transaction. So all these application requires high level of security. Security is the major challenge today's online world. Phishing[1] is an attack which mainly aims for financial gain.

In the year 2012-2013, 37.3 million users around the world were affected to phishing attacks, which is being growing by 87% from 2011-2012. Most often, phishing attacks are targeting users in Russia, US, India, Vietnam and UK. Phishing attacks were most frequently launched from the US, UK, Russia, Germany and India. Google, Yahoo!, Amazon and Facebook are top targets of malevolent users. Online payment systems, online game services, and the bank websites and other financial and credit organizations are also some of the common targets. Over 20% of all attacks banks and other credit and financial organizations. The number of distinct sources of attacks in 2012 and 2013 increased 3.3 Times. More than one half (56.1%) of all identified sources of phishing attacks were located in just 10 countries. In 2012-2013, 1,02,100 Internet users around the world were subjected to phishing attacks every day. This is double the amount of intended victims over the previous period. More than 50% of the total number of individual targets (921 names out of 1,739 in the KSN database) was fake copies of the websites of banks and other credit and financial organizations. Phishing has some local accents: phisher targets are different from country to country, depending on the popularity of local online resources.

So to avoid all these threat we are implementing a technique "Anti-phishing based on Visual Cryptography". In this approach website cross verifies its own identity and proves that it is a genuine website before the end users and make the both the sides of the system secure as well as an authenticated one.

## II. LITERATURE SURVEY

Forged web pages that are created by malicious people to mimic Web pages of real web sites and stealing credential information such as password, user-id, and credit card details of victim/user are called as phishing website. There are a lot of fake phishing websites created and uploaded online every day, luring a number of customers. To prevent from such websites, there is a lot of work that has been done in order to curb the phishing attack. Researchers have been done many works in this area to authenticate the user. Automated challenge Response method ensures two way authentication and simplicity. It is performed using mutual authentication handshake in both directions. So the server ensures that the client knows the secret, and the client also ensures that the server knows the secret, which protects the user against a forged server impersonating the real server. This method also prevents from man in middle attack. But still it has many flaws. This requires higher customer support cost because users often forget the correct answer to a challenge response question and must seek the support from the help desk for a locked user ID to authenticate into the system. Users need to remember multiple piece of information, such as answers to multiple secret questions.

There is DNS-based anti-phishing approach, which includes blacklists, heuristic detection and the page similarity assessment. In this approach hackers tamper with a host's domain name system (DNS) so that requests for URLs or name service return a bogus address and subsequent communications are directed to a phishing or fake site. Blacklist is collection of known phishing Web sites/addresses which are published by authorized entities like Google's and Microsoft's black list. It requires both a client and a server component. Every URL in the blacklist has been verified by the administrator hence the false alarm probability is very low. Heuristic-based anti-phishing technique estimates the Phishing Heuristic Characteristic (like checking URL and Host name) of Webpages. Combining URL-based and HTML-based heuristics is effective to differentiate original site from phishing sites. There are 20 heuristics to evaluate the legitimacy of a website and determine the decisive heuristics. It is easy for the attacker to use technical means to avoid the heuristic characteristics detection. Similarity based approaches to detect phishing web sites are CANTINA (A content-based approach for detecting phishing web sites). It works as follows: Given a web page, calculate the TF-IDF scores of each term on that web page. Generate a lexical signature by taking the five terms with highest TF-IDF weights.

Feed this lexical signature to a search engine, which in our case is Google. If the domain name of the current web page matches the domain name of the N top search results, consider it to be a legitimate web site. Otherwise, consider it a phishing site. It has some drawbacks, it does not include a dictionary for languages other than English. CANTINA suffers from performance problems due to the time lag involved in querying Google. This technique is time-consuming. It needs too long time to calculate a pair of pages, and there is low accuracy rate for this method depends on many factors, such as the text, images and similarity measurement technique. In Page visual similarity assessment approach, web page is personating like real web pages. URL similarity assessment approach, if an URL is similar to a bank's URL, but it is not the bank's URL, it is considered a phishing website's URL. But the speed of calculating the visual similarity between pages is too slow, so it is only used for phishing- spam detection generally.

Some of Offline phishing detecting methods are, Largescale Anti-phishing by Retrospective data- exploration (LARX), all of LARX's phishing filtering operations are based on a cloud computing platform and work in parallel. It used two Amazon Web Services and Eucalyptus as cloud platforms. A physical server is also used for comparison. An OpenID account has a fixed password and several temporary passwords. The fixed password can only be used in bound PCs, that is, we must bind the fixed password on several known PC. Users can login on any PC with a temporary password. It removes the necessity for computer illiterate to learn more so creates a large pool of computer illiterate users, which are vulnerable to attack.

Fuzzy Data Mining(DM) Techniques can be an effective tool in assessing and identifying phishing websites for e-banking since it offers a more natural way of dealing with quality factors rather than exact values. The proposed model is based on Fuzzy Logic(FL) combined with Data Mining algorithms to characterize the e-banking phishing website factors and to investigate its techniques by classifying there phishing types and defining six e-banking phishing website attack criteria's with a layer structure. Single rule for phishing detection like in case of URL is far from enough, so it need multiple rule set for only one type of URL based phishing detection. Textual and visual content based anti-phishing mechanism takes into account textual and visual contents to measure the similarity between the protected web page and suspicious web pages. This is required in the classifier for determining the class of the web page and identifying whether the web page is phishing or not.

But these all systems have their own drawbacks like LARX showed the insignificant trivial influence of the (Page Style & Content) criteria along with (Social Human Factor) criteria in the phishing detection final rate result. And Mutual authentication method problems like hijacking account setup, theft of the trusted device and attacks on the network.

## III. VISUAL CRYPTOGRAPHY

Security is the major challenge in today's online world. So to provide security mostly used technique is cryptography. Cryptography technique is used to exchange the message between sender and receiver in coded and secret form which is done by encryption by the sender and decryption only by the intended receiver.

Visual Cryptography[2],[3] is a cryptographic technique which allows visual information(for example Image) to be encrypted in such a way that decryption does not require complex mathematical computation instead can be done by human visual system. Visual Cryptography Scheme(VCS) allows secret sharing of images without cryptographic computations. This technique was discovered by Shamir and Naor. This can be achieved by following schemes:

**1) (2,2) Threshold VCS scheme**: In this scheme, the message (image) is taken and is encrypted into two shares and for decryption both the shares are stacked together and image is reconstructed.

**2) (n,n) Threshold VCS scheme:** In this scheme, the message (image) is taken and is encrypted into n shares and for decryption this n shares are stacked together and image is reconstructed.

**3) (k,n) Threshold VCS scheme*:*** In this scheme, the message (image) is taken and is encrypted into n shares and for decryption only k shares are stacked together and image is reconstructed.

We are implementing (2,2) VCS for our system. In (2,2) secret image sharing scheme each image is divided into two shares and the original image cannot be reconstructed from3 any one of the either shares. In this, each pixel P of the image is encrypted into two sub pixels. These sub pixels are called as shares. Fig.1 depicts the shares of black pixel and white pixel.

The choice of white and black pixel can be determined randomly and each pixel has two choices as shown in fig.1.

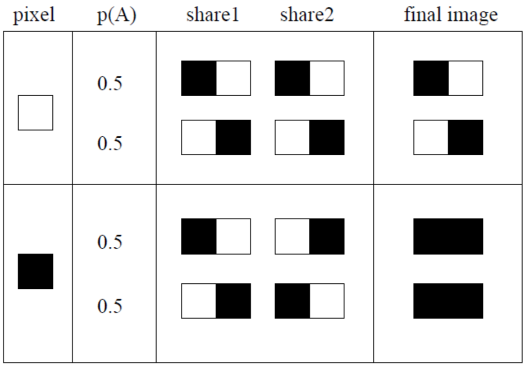

Fig. 1 (2,2)- Threshold VCS scheme

To perform decryption process, these two shares are superimposed or stack together such that we recover back the original image. 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 i.e. grey pixel.

After performing the decryption process human visual system can recognize the reconstructed image.

### *A. Current Methodology*

In present scenario, if the user wants to do online transaction or access confidential information then he gets login to the bank account or any secure mail account. For doing this user enters information like user id, password, account number etc., on login page. The attacker can hack the confidential information and then redirect the user to the original website. User will be unknown to this attack. Phisher can use this confidential information for fraudulent activities. This is shown in fig.2.

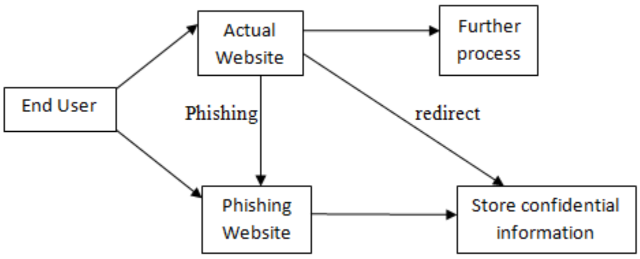

Fig. 2 Current scenario

### *B. Proposed Methodology*

For phishing detection and prevention, the proposed new methodology overcomes the drawback of previous anti-phishing techniques. The methodology is based on the Anti-Phishing[3],[4] Image CAPTCHA Validation scheme using Visual Cryptography. Using this methodology password and other confidential information is prevented from phishing attack.

The proposed approach is divided into two phases:

- Registration Phase
- Login Phase

1) **Registration Phase:** In the registration phase user is asked to enter details. User enters password and also enters key string at the time of registration, which may be the combination of alphabets and numbers. Now this entered key string is concatenated with randomly generated key string and an image CAPTCHA is formed. The generated image CAPTCHA is divided into two shares using (2, 2) VCS algorithm. One of the shares is kept with the user and other share is stored in server database securely. The original image CAPTCHA and share is provided to user for later verification during login phase. After the registration the user may change the key string as and when required. Registration phase is shown in fig.3.

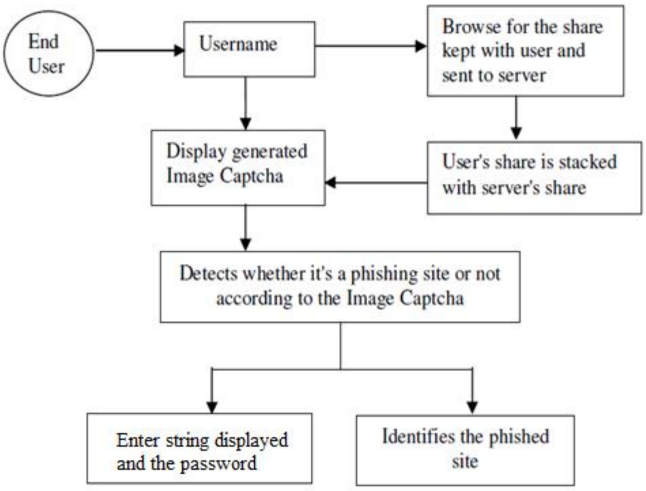
End User
Username
Browse for the share kept with user and sent to server
Display generated Image Captcha
User's share is stacked with server's share
Detects whether it's a phishing site or not according to the Image Captcha
Enter string displayed and the password
Identifies the phished site

Fig. 3 During Registration Phase

2) **Login Phase:** In login phase first the user is asked to enter user id. The user is also asked to enter the share which is kept which was provided at the time of registration. Now, this share is sent to the server then server searches the other corresponding share and stacks it with user share to produce the image CAPTCHA. The reconstructed image CAPTCHA is displayed to the user. The end user now needs to match the displayed reconstructed image CAPTCHA with the image CAPTCHA generated at the time of the registration. The end user needs to enter this matched image CAPTCHA text so that the server can check whether the user is human or not. User is also asked to enter the password to check whether the user is authorized user. If the displayed image CAPTCHA does not match or there is no image CAPTCHA displayed then the site is phishing website. Login phase is shown in fig.4.

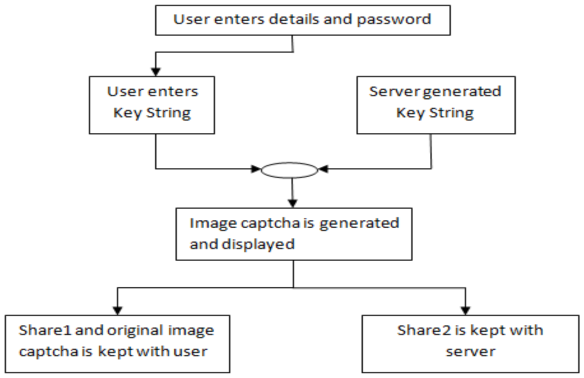
User enters details and password
User enters Key String
Server generated Key String
Image captcha is generated and displayed
Share1 and original image captcha is kept with user
Share2 is kept with server

Fig. 4 During Login Phase

### C. VCS 2×2 Algorithm

**Algorithm 1** VCS 2by2 algorithm

```
procedure GENERATEKEYIMAGE($w, h$)
    $w \leftarrow w * 2$                ▷ w= width of the original image
    $h \leftarrow h * 2$                ▷ h= height of the original image
    $n \leftarrow w * h$                ▷ n = no. of pixels in the original image

    Each 2x2-pixel-pack has 2 randomly set pixels
    $y \leftarrow 0$
    for $y \leftarrow 0, h$ do
        $x \leftarrow 0$
        for $x \leftarrow 0, w$ do
                                        ▷ determine the two pixels
            $px1 \leftarrow random\ number\ between\ [0-4]$
            $px2 \leftarrow random\ number\ between\ [0-4]$
            while $px1 \equiv px2$ do
                $px2 \leftarrow random\ number\ between\ [0-4]$
            end while
            $px1x \leftarrow (px1 < 2)?px1 : px1 - 2;$
            $px1y \leftarrow (px1 < 2)?0 : 1;$
            $px2x \leftarrow (px2 < 2)?px2 : px2 - 2;$
            $px2y \leftarrow (px2 < 2)?0 : 1;$
            fill rectangle ($x$+$px1x$, $y$+$px1y$, 1, 1) with black colour
            fill rectangle ($x$+$px2x$, $y$+$px2y$, 1, 1) with black colour
            $x \leftarrow x + 1$
        end for
        $y \leftarrow y + 1$
    end for
    return $2(n * m)$
                                        ▷ For each pixel in the
original image, 2m sub pixels are written, and each share contains
n * m pixels as such, a total of 2(n * m) pixels are written in the
encryption process,where m=no. of sub-pixels
end procedure
```

This algorithm is used for the encryption process. In the encryption process the 2 by 2 Visual Cryptography Scheme [5] uses two transparent images. One image contains random pixels and the other image contains the secret information. The algorithm 1, below is used in to generate key, which is the used for encryption process. The algorithm 2, is used for encryption. Here, the algorithm takes key image and CAPTCHA image as an argument and using these the encrypted image is generated.

In the decryption process, the two shares created during encryption process are stacked together which results is reconstructed CAPTCHA image. This is displayed to the user and hence the website proves that it is genuine website. Phisher cannot provide the correct CAPTCHA image to the user and then the user detects that it is phished website.

**Algorithm 2** Image Encryption Algorithm

```
procedure ENCRYPTIMAGE($KImg, SImg$)
                        ▷ where KImg is KeyImage,SImg is SourceImage
                                        ▷ check for key/source file match
    if ($SImg.Width \neq KImg.Width/2$)||($SImg.Height \neq KImg.Height/2$) then
        return $error$
    end if
    Resize the SImg to the size of the KImg.
    Create blank EImg of size KImg
                                        ▷ where EImg is EncryptedImage
    Fill it with a fully transparent white

    Each 2x2-pixel-pack has 2 pixels to set
    $y \leftarrow 0$
    for $y \leftarrow 0, imgEncr.Height$ do
        $x \leftarrow 0$
        for $x \leftarrow 0, imgEncr.Width$ do
▷ because 1 black pixel of the original image is now a square of
4 black pixels ▷ only the first pixel has to be checked ▷ where
RGB=getRGBvalue
            if ($SImg.RGB(x, y) \equiv Color.BLACK.RGB()$)
then
  ▷ Write the two pixels to complete the block together with the
key
                if ($KImg.RGB(x, y) >>> 24 \equiv 0$) then
                    fill $EImg$ ($x, y, 1, 1$) with black colour
                end if
                if ($KImg.RGB(x + 1, y) >>> 24 \equiv 0$) then
                    fill $EImg$ ($x + 1, y, 1, 1$) with black colour
                end if
                if ($KImg.RGB(x, y + 1) >>> 24 \equiv 0$) then
                    fill $EImg$ ($x, y + 1, 1, 1$) with black colour
                end if
                if ($KImg.RGB(x + 1, y + 1) >>> 24 \equiv 0$) then
                    fill $EImg$ ($x+1, y+1, 1, 1$) with black colour
                end if
            else
                if ($KImg.RGB(x, y) \equiv 0$) then
                    fill $EImg$ ($x, y, 1, 1$) with black colour
                end if
                if ($KImg.RGB(x + 1, y) \equiv 0$) then
                    fill $EImg$ ($x + 1, y, 1, 1$) with black colour
                end if
                if ($KImg.RGB(x, y + 1) \equiv 0$) then
                    fill $EImg$ ($x, y + 1, 1, 1$) with black colour
                end if
                if ($KImg.RGB(x + 1, y + 1) \equiv 0$) then
                    fill $EImg$ ($x+1, y+1, 1, 1$) with black colour
                end if
            end if
            $x \leftarrow x + 1$
        end for
        $y \leftarrow y + 1$
    end for
    return $EImg$

end procedure
```

### *D. Analysis*

To encrypt an image using the 2 out of 2 algorithm[2], each pixel in the original image must be read, and a block of mm sub-pixels must then be written to each share. Thus, for each pixel in the original image, 2*m* sub pixels are written, and each share contains *n***m* pixels. As such, a total of 2(*n***m*) pixels are written in the encryption process. As long as *m*<*n*, we thus have a linear time complexity of O(*n*) for the encryption algorithm. When superimposition is being done, each sub-pixel in each share must be read sequentially, computing the Boolean OR of the sub-pixels from each share as they are read. This computation requires O(*m*) time, and there are *n* such computations. Once again, as long as *m*<*n*, decryption takes place in linear time. Unlike previous research of Anti-phishing technique, we are providing one more password field. Even if the phisher somehow gets the user's share he would not have the password of user he has entered while registration. So, the phishing attack is prevented. Figure 5 shows the result of creation and stacking of shares.

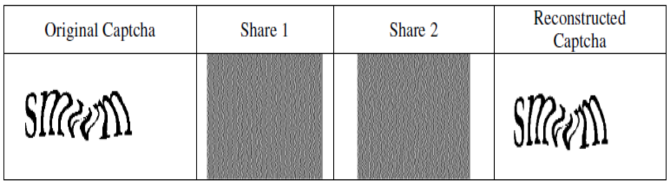

Fig. 5 Creation and stacking of shares

## IV. CONCLUSION

As we have implemented anti-phishing technique based on visual cryptography, the complex mathematical calculations are reduced to large extent unlike other encryption algorithms. It also detects that whether the user is human or not, as only human being can recognize the decrypted CAPTCHA image. The website first proves that it is a genuine website before the user gets logged in. The user also has to prove that he is authorized user by entering his user id, share1 the displayed CAPTCHA and also a password. Thus it makes both the side of the system secure by detecting and preventing phishing attacks.

## REFERENCES

[1] Ollmann G., *The Phishing Guide Understanding and Preventing Phishing Attacks*, NGS Software Insight Security Research.

[2] Moni Naor and Adi Shamir, *Visual Cryptography*, advances in cryptology-Eurocrypt, pp 1-12,1995.

[3] M. Naor and A. Shamir, *Visual Cryptography*, in Proc. EUROCRYPT,1994, pp. 1-12.

[4] Chuan Yue and Haining Wang, *Anti-Phishing in Offense and Defense*, The College of William and Mary.

[5] Divya James and Mintu Philip, *A Novel Anti Phishing Framework Based On Visual Cryptography*, International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.1, January 2012

[6] Rohith S, Vinay G, *A Novel Two Stage Binary Image Security System Using (2,2) Visual Cryptography Scheme*, IJCER, 2012, Page 642.

[7] Chandramathi S., Ramesh Kumar R., *An Overview Of Visual Cryptography*, International Journal of Computational Intelligence Techniques, Vol.