

New Antiphishing Approach Based on Probabilistic Model of (t, n) VC Scheme for Commerce Bank

Ms. Mangala S. Wale
Dept. of Computer Engineering
Jspm's BSOTR(w), Wagholi
Pune, India

Asst. Prof. Gayatri M. Bhandari
Dept. of Computer Engineering
Jspm's BSOTR(w), Wagholi
Pune, India

Abstract - Phishing is one of the attacks that became popular recently. It is a duplicate copy of web page used to acquire personal and confidential information of user like name, id, email, contacts, password and credit card details by legitimate entity in an electronic communication. In this project we have proposing a new technique named as "New Antiphishing Approach Based On Probabilistic Model Of (t, n) VC Scheme For Commerce Bank" to solve the problem of phishing. Image captcha is used for authentication based on (t, n) VC scheme. Visual cryptography is to encrypt a secret image into some shares (transparencies) such that any qualified subset of the shares can recover the secret visually. The (t, n) is VC scheme required t number of shares (transparencies) out of n number of shares (transparencies) The stacking of $t-1$ or fewer shares(transparencies) is unable to extract any information about the secret. To reduce the overhead of generating n shares (transparencies) of user, using (t, n) VC scheme with unlimited n . Here original image captcha is divided into n number of shares, where one share stores with user and the remaining shares store in server. User's share is stacked with servers share to reveal the secret image for identifying phishing website; the individual sheet images do not reveal the secret image. The decryption is possible only when commerce bank can provide both shared at a time. In this project we used (t, n) VC Scheme based on basis matrices and the probabilistic model where optimal contrast with $t=2$ to 6 and $n=2$ to 10 are solved in this project.

Index Terms – Contrast, Image Captcha, Phishing, Visual Cryptography(VC) scheme, Security

I. INTRODUCTION

Now a day's online transaction are very common, also online purchases and online account management has become very common. There are various attacks with which the hacker can hack the user accounts and credentials like password, name, id, email, contacts, and credit card information, etc. In these types of attacks, phishing is a major security, where Phishers target to customers of banks and online payment services.

Hackers can steal sensitive information such as Online banking passwords and credit card information from users by using various phishing techniques. In this kind of attack Phishers attempt to fraudulently acquire sensitive information by masquerading as a trustworthy person or business in an electronic communications.

Here, we used the concept of an improved visual cryptography and Image Processing. In Image Processing input image converted into improved form of the same

image as output image. Visual cryptography scheme is like to image processing where image is encoded and send to the receiver side for decryption to get the original image [6]. Decryption is done by human visual system, so user needs to remember original image for identifying phishing website. VCS is used to divide the image captcha into shares and in order to reveal the original image captcha appropriate number of shares should be combined. In this project, secret image is encoded into shares, according to Visual Cryptography scheme. Using $(2, 2)$ VCS image captcha is decomposed into two shadow images called as shares or transparencies. One secret dynamic captcha share stored in server side and other secret dynamic captcha share sent to the user. When recompose the original image captcha these 2 out of 2 shares needed to combine reveals the secret image. Using (t, n) VCS, any t out of n shares needed to combine reveals the secret image (dynamic Image captcha). One stored in user database and remaining stored in server databases. When recompose the original image captcha these 2 out of n shares must be combined together reveals the secret image [10].

The probabilistic model of the Visual Cryptography scheme [10] is based on the basis matrices, but only one column of the matrices is chosen to encode a binary secret pixel rather than the traditional VC scheme utilizing the whole basis matrices. The size of the generated transparencies is identical to the secret image. Yang [11] also proposed a probabilistic model of (t, n) VC scheme, and the two cases $(2, n)$ and (n, n) are explicitly constructed to achieve the optimal contrast. Based on Yang [11], proposed a generalized VC scheme in which the pixel expansion is between the probabilistic model of VC scheme and the traditional VC scheme. Contrast is one of the important performance metrics for VC schemes. Generally, the stacking revelation of the secret with higher contrast, represents the best visual quality, and therefore the stacking secret with high contrast is the goal of pursuing in VC designs [8]. This project designs the implementation of (t, ∞) based on the probabilistic model. Algorithm 1 and 2, with computationally feasible operations. We derive optimization problem $L(t)$ to solve the maximal contrast of the proposed (t, ∞) VC scheme. Table 1 shows the notations and meaning used in this project.

Notations	Meaning
t	The threshold of a (t, n) VC scheme
n	The number of generating share (transparencies) of a (t, n)
m	The number of sub pixels to encode a secret Pixel; i.e., the width of the basis matrices.
X,Y	Two vectors to record the nonzero terms in Δp
L(t)	An optimization problem to find the optimal Contrast of (t, $n \rightarrow \infty$) VC scheme.
S	A binary secret image.
s	A ready-to-process pixel taken from S
T_i	The ith generated transparency.
t_i	A pixel at T_i , and the position corresponds to the position of s.
n'	To specify number of transparencies (shares).
n''	To specify next number of transparencies (shares).
Z	An index table $Z[w,h]$ where is the index of the used memoryless sequence $E(xZ[w,h])$ to encode the secret pixels [w, h].
p	Probability function

TABLE 1: Notations and meaning

II. RELATED WORK

Automated Challenge Response Method [1] gives the concept of Anti phishing technique using various methods. Phishing is a combination of social engineering and technical deception to steal consumer's personal identity data and financial account credentials. Even though there are numerous methods reported to avoid Phishing each method has its own limitations. This paper addresses one of the limitations in Transaction Authentication Number method. One such authentication mechanisms, includes challenge generation module from the server, which in turn interacts with the Challenge - Response interface in the client and request for response from user. A challenge - Response module in turn will call the get response application which is installed on the client machine. The method ensures two way authentications and simplicity. It prevents man-in-the middle attacks.

DNS - based Antiphishing approach technique which mainly includes blacklists, heuristic detection, the page similarity assessment. But they do have some shortcomings. The blacklist is a DNS based anti-phishing approach technique now most commonly used by the browser [2]. Heuristic-based anti-phishing technique is to estimate whether a page has some phishing heuristics characteristics. For example, some heuristics characteristics used by the Spoof Guard [3] toolbar include checking the host name, checking the URL for common spoofing techniques, and checking against previously seen images. The following technologies used, but they have several drawbacks:

1. 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 too short and the establishment of

blacklist has a long lag time, the accuracy of the blacklist is not too high.

2. Heuristic-based anti-phishing technique, with a high probability of false and failed alarm, and it is easy for the attacker to use technical means to avoid the heuristic characteristics detection.
3. Similarity assessment based technique [2] is time-consuming. It needs too long time to calculate a pair of pages, so using the method to detect phishing websites on the client terminal is not suitable. And there is low accuracy rate for this method depends on many factors, such as the text, images, and similarity measurement technique. However, this technique (in particular, image similarity identification technique) is not perfect enough yet.
4. An offline phishing detection system [4] named LARX, an acronym for Large-scale Anti-phishing by Retrospective data exploration to counter phishing attacks has been proposed, but ends out to be offline and not dynamic to real world events.
5. An Antiphishing mechanism [5] using Bayesian approach happens to be better. This framework synthesizes multiple cues, i.e., textual content and visual content, from the given web page and automatically reports a phishing web page. But it also happens to be computationally expensive processed.

Naor and Shamir [8] introduced the visual cryptography scheme (VCS) as a simple and secure way to allow the secret sharing of images without any cryptographic computations. Visual cryptography schemes were independently introduced by Shamir and Blakley, and their original motivation was to safeguard cryptographic keys from loss. These schemes also have been widely employed in the construction of several types of cryptographic protocols and consequently, they have many applications in different areas such as access control, opening a bank vault, opening a safety deposit box, or even launching of missiles. Naor and Shamir [8] proposed a (t, n) threshold VC scheme based on basis matrices, and the model had been further studied and extended. It defines a contrast formula which has been widely used in many studies. Based on the definition of contrast, there are studies attempting to achieve the contrast bound of VCS.

III. METHODOLOGY

Software Requirement

- Install the apache-tomcat-7.0.5
- Install the eclipse IDE
- Install the MySql Server
- Install MysqlGuiTool
- Windows XP, Windows 7

Hardware Requirements

- Pentium 4 Processor
- Hard disk 1 GB
- Ram 512

A. Basis Matrices

As proposed the basis matrices of VC scheme by Naor, a white-and-black secret image or pixel is also described as a binary image or pixel. In the basis matrices, to encode a binary secret image, each secret pixel white black will be turned into blocks at the corresponding position of transparencies, respectively. Each block consists of subpixels and each subpixel is opaque or transparent. Throughout this paper, we use 0 to indicate a transparent subpixel and 1 to indicate an opaque subpixel. If any two subpixels are stacked with matching positions, the representation of a stacked pixel may be transparent, when the two corresponding pixels are both transparent. Otherwise, the stacked pixel is opaque.

Let \oplus denote the stacking operation, defined as $0 \oplus 0 = 0, 0 \oplus 1 = 1, 1 \oplus 0 = 1, 1 \oplus 1 = 1$. Actually, \oplus can be treated as the bitwise operation "OR." It is noted that we use the notation to indicate the stacking of the two transparencies and since and can be treated as two Boolean matrices.

B. (2, 2) Visual Cryptography Scheme(VCS)

One of the best known techniques to protect data is cryptography. It is the art of sending and receiving encrypted messages that can be decrypted only by the sender or the receiver. Encryption and decryption are accomplished by using mathematical algorithms in such a way that no one but the intended recipient can decrypt and read the message. VCS is a cryptographic technique that allows for the encryption of visual information such that decryption can be performed using the human visual system. We can achieve this by one of the following access structure schemes [11].

The (2, 2) Threshold VCS Scheme: This is a simplest threshold scheme that takes a secretMessage and encrypts it in two different shares that reveal the secret image when they are overlaid.

In (2,2) VCS the first 2 represents the minimum number of share images needed to recover a secret image. The second 2 represents the total number of share images produced. The VCS model is dependent on the basis matrix which forms the entire model. In linear algebra basis is a set of linearly independent vectors, can represent every vector in the vector space. The entire model of (2, 2) VCS can be described by tow basis matrices one for a black pixel and one for a white pixel. The basis matrices of (2,2) VCS are:

$B_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

In a basis matrix element 1 means the presence of a black pixel in the share image generated from this matrix and element 0 means the presence of a white pixel. The rows of a basis matrix correspond to the share images and describe how the pixel in secret image is divided in share image. For example consider the pixel to be shared is black pixel, and then the dealer takes the B1 basis matrix and examines the rows. For the share1 image he copies the black pixel as a combination of black pixel and white pixel as in 1st row of B1 matrix. For the share2 image he copies black pixel as a white and black pixel combination as in 2nd row of the B1 matrix. In the same way for the white pixel, share1

image gets the pixel as in 1st row of B0 matrix and share2 image gets the pixel as in 2nd row of B0 matrix. Note that the choice of shares for a white and black pixel is randomly determined (there are two choices available for each pixel). Neither share provides any clue about the original pixel since different pixels in the secret image will be encrypted using independent random choices [9]. When the two shares are superimposed, the value of the original pixel P can be determined. 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 as shown in fig 1.

Pixel	Probability	Shares		Superposition of the two shares	
		#1	#2		
□	$p = 0.5$	■	■	■	White Pixels
	$p = 0.5$	□	□	□	
■	$p = 0.5$	■	□	■	Black Pixels
	$p = 0.5$	□	■	■	

Fig 1. (2, 2) visual cryptography scheme.

C. (t, n) Share Generation Visual Cryptography Scheme

In this project, we propose a probabilistic model of (t, n) VC scheme with unlimited n. The proposed scheme allows changes of users without regeneration and redistribution of VC transparencies, which reduce the computing and communication resources in accommodating user changes. The scheme is capable of generating an arbitrary number of transparencies and the explicit algorithms are proposed to generate the transparencies [10]. For a group with n' initial users, the proposed Algorithm 1 explicitly generates the required n' transparencies. For n'' newly joining participants, the n'' new transparencies can be explicitly generated through Algorithm 2, and the new transparencies can be distributed to the new participants without the need to update the original transparencies. The secondary contribution is that this paper designs an (t, ∞) implementation of VC based on the probabilistic model, and the proposed scheme allows the unlimited number of users. For the conventional VC scheme, the mathematical manipulations of infinite size of basis matrices and variables are often required, which is computationally prohibitive. We also derive an optimization problem L(t) to solve the maximal contrast of the proposed VC scheme.

t	X and Y
2	$Y = (1/2, -1, 1/2)$ $X = (0, 1/2, 1)$
3	$Y = (-1/3, 2/3, -2/3, 1/3)$ $X = (0, 1/4, 3/4, 1)$
4	$Y = (1/4, -1/2, 1/2, -1/2, 1/4)$ $X = (0, (2-)/4, 1/2, (2+)/4, 1)$
5	$Y = (-1/5, 2/5, -2/5, 2/5, -2/5, 1/5)$ $X = (0, (3-)/8, (5-)/8, (3+)/8, (5+)/8, 1)$
6	$Y = (1/6, -1/3, 1/3, -1/3, 1/3, -1/3, 1/6)$ $X = (0, (2-)/4, 1/4, 1/2, 3/4, (2+)/4, 1)$

TABLE II: (X, Y) of some (t, ∞) schemes for 2 ≤ t ≤ 6

For a given value of t , the transparencies can be continuously generated with the (t, ∞) OptPrVC scheme. However, practical applications require the algorithm to terminate within finite steps. To meet the requirement, a finite number n' is used to specify the number of transparencies in the algorithm. The algorithm requires (X, Y) , obtained by solving $L(t)$ [10]. The outputs of Algorithm 1 are transparencies and an index table Z , where $Z[w, h]$ is the index of the used memory less sequence $E(xZ[w, h])$ to encode the secret pixel $s[w, h]$.

Algorithm 1. The algorithm of (t, ∞) OptPrVC scheme
Input: A binary secret image S , two positive integers t, n' , and two vectors (X, Y) .
Output: n' transparencies T_1, T_2, \dots an index table Z .
 1 for each pixel $s[w, h]$ in S do
 2 if $s[w, h] = \text{white}$ then
 3 Generate an integer $z \in \{t-2k \mid k=0, 1, \dots, [t/2]\}$ and $P(z=t-2k) =$
 4 else
 5 Generate an integer $z \in \{t-1-2k \mid k=0, 1, \dots, [t-1/2]\}$ and $P(z=t-1-2k) =$
 6 end if
 7 $Z[w, h] = z$.
 8 for $k=1$ to n' do
 9 Assign randomly $T_0 k [w; h]$ to 0 or 1 where $P(T_0 k [w; h] = 0) = xz$.
 10 end for
 11 end for

D. (t, ∞) Next Share Generation Visual Cryptography Scheme Algorithm

In the first round, we use Algorithm 1 to generate transparencies and Z . If we need not to generate more transparencies in the future, Z is not required and discarded. Otherwise, has to be stored in a safe place, and we can generate more transparencies $T_1', T_2', \dots, T_{n'}'$ by utilizing Z .

Algorithm 2. The algorithm of (t, ∞) OptPrVC scheme by the index table Z .
Input : An index table Z , a positive integer n'' , and a vector X .
Output: n'' transparencies $T_1'', T_2'', \dots, T_{n''}''$.
 1 for each $Z[w, h]$ in Z do
 2 for $k=1$ to n'' do
 3 Assign randomly $T_0 k [w; h]$ to 0 or 1 where $P(T_0 k [w; h] = 0) = xz$
 4 end for
Send for

IV. DESIGN PROCESS

The Antiphishing using VCS technique consists of 3 phases registration phase, login phase and Share recovery phase[6].

- A. Registration phase
- B. Login phase
- C. Share recovery phase

A. Registration phase

In the registration phase, a key string (password) is asked from the user at the time of registration for the secure website. The key string can be a combination of alphabets and numbers to provide a more secure environment. This string is concatenated with randomly generated string in the server and an image captcha is generated. The image captcha is divided into two shares such that one of the shares is kept by the user and the other share is kept in the server. The users share and the original image captcha are sent to the user for later verification during the login phase. The image captcha is also stored in the actual database of any confidential website as confidential data. After the registration, the user can change the key string when it is needed. The registration process is depicted in Fig 2.

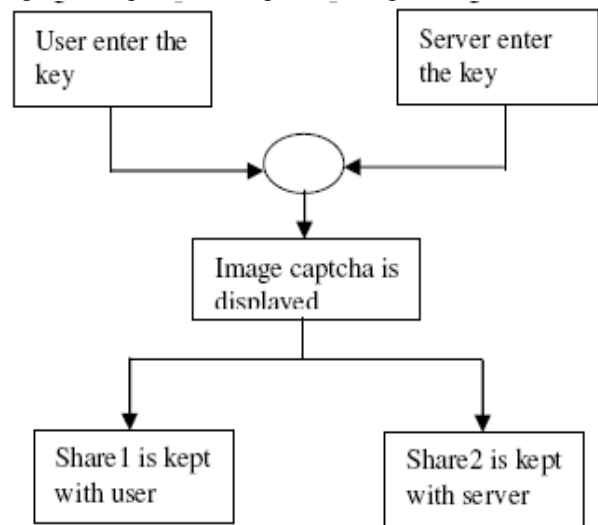


Fig 2. Registration Phase

Fig.3 shows the resulting page of the registration phase where the user is going to enter his details when he is going to register.

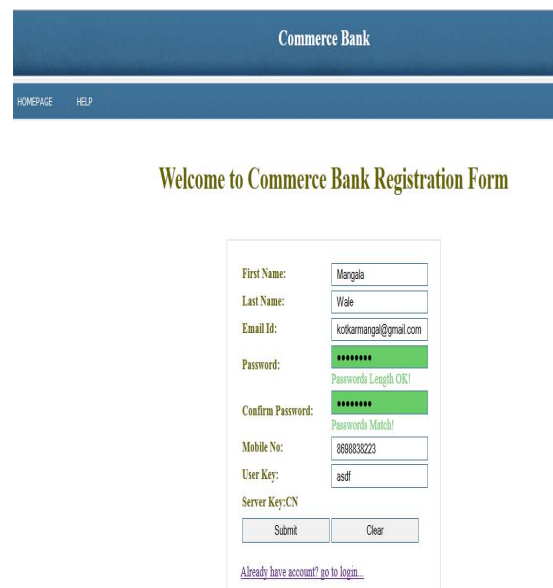


Fig.3: Resulting page of the registration phase

B. Login Phase

In the Login phase first the user is entered Email id. Then the user is asked to enter his share which is kept with him. This share is sent to the server where the users share and share which is stored in the database of the website, for each user, is stacked together to produce the image captcha. The image captcha is displayed to the user. Here the end user can check whether the displayed image captcha [7] matches with the captcha created at the time of registration. The end user is required to enter the text displayed in the image captcha and this can serve the purpose of password and using this, the user can log in into the website. Using the username and image captcha generated by stacking two shares one can verify whether the website is genuine/secure web site or a phishing website and can also verify whether the user is a human user or not. Fig 4 can be used to illustrate the login phase.

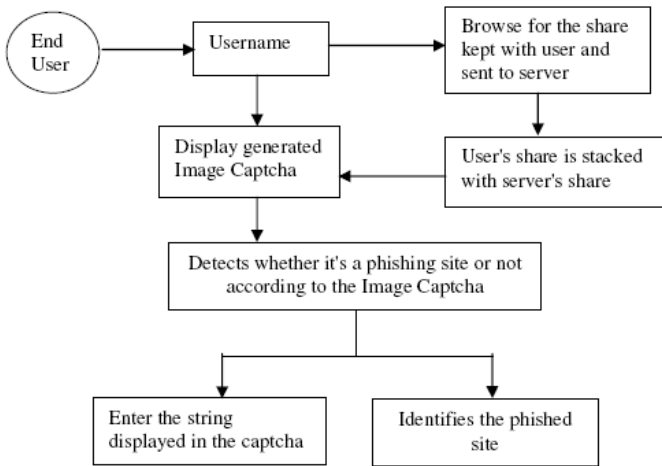


Fig 4. Login Phase

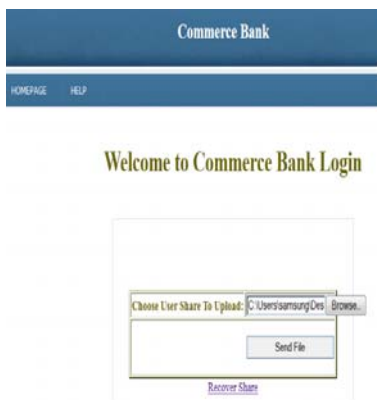


Fig.5 Resulting page of the Login phase

C. Share recovery phase

The Share recovery Phase is used when user lost or corrupts his share. In the registration Phase when the User enters username and try upload his share from the server. If she/he lost or corrupt his/her share then he request for new share at that time server crosscheck whether the user is

authorized or not. The server uses next share algorithm (t, ∞) for generating new share, which is compatible with user's share [10]. The server generates a new share for the user. Users download new share and process continue with login page is as shown in Fig 6.

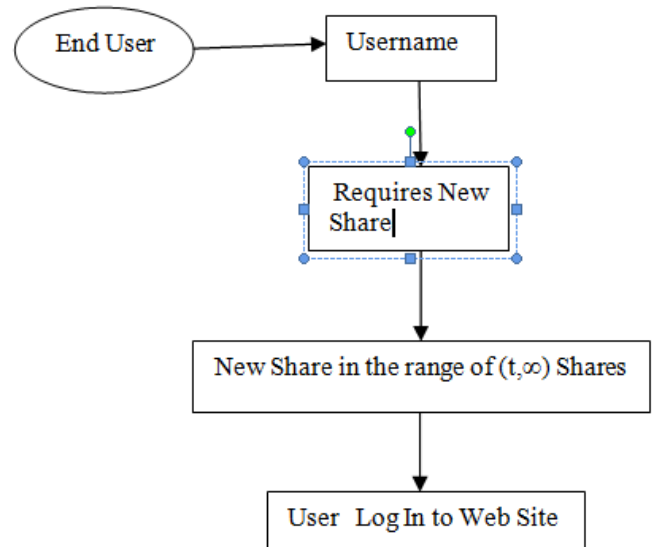


Fig. 6. Share Recovery phase

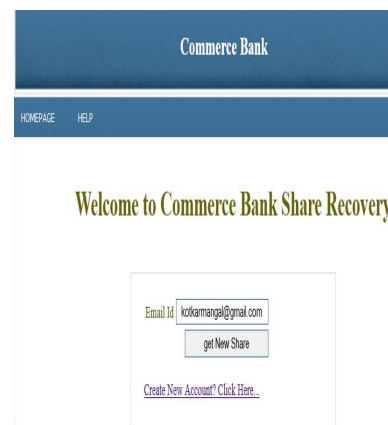
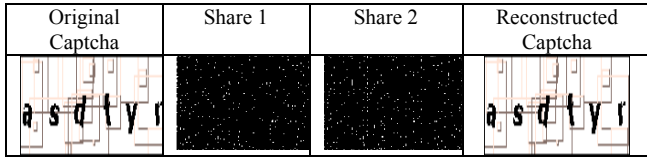


Fig 7. Resulting page of share recovery page

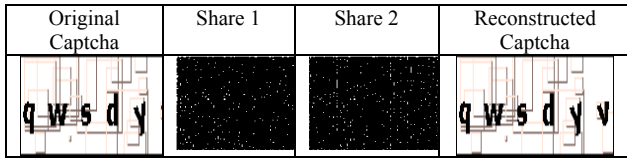
V. EXPERIMENTAL RESULT

The shares generated are combination of black & white pixels whose stacking results in original image which is also a black & white image. Fig. 8 shows the result of creation and stacking of shares. In the registration phase the most important part is the creation of shares from the image captcha where one share is kept with the user and other share can be kept with the server. For login, the user needs to enter a valid mail id in the given field. Then he has to browse his share and process. At the server side the user's share is combined with the share in the server and an image captcha is generated. The user has to enter the text from the image captcha in the required field in order to login into the website.

Case 1:



Case 2:



Case 3:

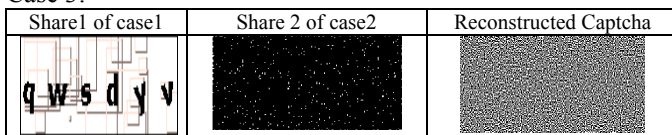


Fig. 8. Creation and stacking of shares

The entire process is depicted in Fig.5 as different cases. Case 1 and Case 2 illustrates the creation and stacking of shares of two image captchas resulting in original captcha. In Case 3 share 1 of first image captcha is combined with share 2 of second captcha resulting in unrecognizable form of captcha.



Fig 9. Binary Secret Image

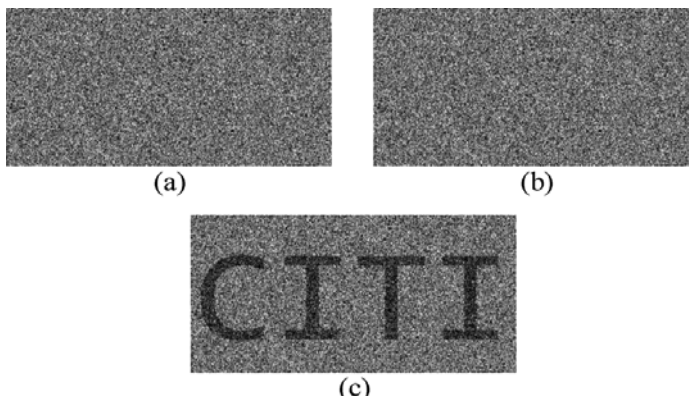


Fig. 10. Results of $(2, \infty)$ OptPrVC scheme (a) T1. (b) T2 . (c) $T1 \oplus T2$.

Fig 9. Shows the Binary Secret Image S, Fig 10 (a)-(b) shows the experiment for $t=2$. It shows two generated transparencies T1 and T2, The stacking result $T1 \oplus T2$, is as shown in fig 7(c). We observe that the characters on the stacking result are clear.

VI. CONCLUSION

The project verifies whether the website is a secure website or a phishing website. If website is phishing then it can't display the image captcha for that specific user. It validates image Captcha and ensure that the site as well as the user is permitted one or not. So, using image Captcha technique, no machine based user can crack the password or other confidential information of the users. It also prevents intruder's attacks on the user's account. It is useful to prevent the attacks of phishing websites on financial web portal, banking portal, online shopping market. Initially, the proposed scheme is based on basis matrices, but the basis matrices with infinite size cannot be constructed practically. Therefore, the probabilistic model is adopted in the scheme. As the results listed in Table II, the proposed scheme also provides the alternate verification for the lower bound. This project is used for creating new share when user lost his share. We implemented a new (t, ∞) Visual Cryptography algorithm. The algorithm is useful in the sense that if you need one more share, you will get it, no need to perform entire visual cryptography and generation of all shares. Also, execution of algorithm does not need high configuration resources, and it can be easily run with good performance on lower configuration infrastructure.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contributions of M.Naor and A.Shamir for their work in the field of visual cryptography.

REFERENCES

- [1] Thiyagarajan, P.; Venkatesan, V.P.; Aghila, G.; "Anti-Phishing Technique using Automated Challenge Response Method", in Proceedings of IEEE- International Conference on Communications and Computational Intelligence, 2010.
- [2] Sun Bin.; Wen Qiaoyan.; Liang Xiaoying.; "A DNS based Anti-Phishing Approach," in Proceedings of IEEE- Second International Conference on Networks Security, Wireless Communications and Trusted Computing, 2010.
- [3] Nourian, A.; Ishtiaq, S.; Maheswaran, M.; " CASTLE: A social framework for collaborative antiphishing databases", in Proceedings of IEEE- 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing, 2009.
- [4] Sid Stamm, Zulfikar Ramzan, "Drive-By Pharming", v4861 LNCS, p495-506, 2007, Information and Communications Security - 9th International Conference, ICICS 2007, Proceedings.
- [5] Anthony Y. Fu, Liu Wenyin, "Detecting Phishing Web Pages with Visual Similarity Assessment Based on Earth Mover's Distance (EMD)", October/December 2006.
- [6] A Novel Anti Phishing Framework Based on Visual Cryptography, 2012, Divya James, Mintu Philip. 2012.
- [7] Matthew Dailey, Chanathip Namprempre, "A Text-Graphics Character CAPTCHA for Password Authentication".
- [8] M. Naor and A. Shamir, V.Venkateswara Reddy , "Visual cryptography", 1995, vol. 950, LNCS, pp. 112.
- [9] R. Ito, H. Kuwakado, and H. Tanaka, "Image size invariant visual cryptography", IEICE Trans. Fundam. Electron., Commun., Comput. Sci., vol. 82, pp. 2172-2177, Oct. 1999.
- [10] Sian-Jheng Lin and Wei-Ho Chung, "A Probabilistic Model of (t, n) Visual Cryptography Scheme With Dynamic Group", 2012.
- [11] N. Yang, "New visual secret sharing schemes using probabilistic method", Pattern Recognit. Lett., vol. 25, no. 4, pp. 481-494, Mar 2004.