Anti-Phishing System Using Visual Cryptography

Vikas Sahare
Research Scholar, MIT Alandi,Pune,Maharashtra, India.

S.A Jain
Associate Proffesor, MIT Alandi,Pune,Maharashtra, India.

Manish Giri
Assistant Proffesor, MIT Alandi,Pune,Maharashtra, India.

Abstract — With the coming of web, different online assaults has been expanded and among them the most famous assault is phishing. Phishing is an endeavor by an individual or a gathering to get individual private data, for example, passwords, credit card data from clueless exploited people for character theft, financial increase and other false exercises. Fake web sites which seem fundamentally the same to the first ones are being hosted to accomplish this. In this paper we have proposed another approach named as "Anti-phishing system using visual cryptography in cloud" to take care of the issue of phishing. Here an image based confirmation utilizing Visual Cryptography is implemented. The utilization of visual cryptography is investigated to safeguard the privacy of a picture captcha by breaking down the image captcha into two shares (known as sheets) that are put away in separate database servers (one with client and one with server) such that the first picture captcha can be uncovered just when both are all the while accessible; the individual sheet pictures do not uncover the character of the first picture captcha. When the original picture captcha is uncovered to the client it can be utilized as the secret word. Utilizing this site cross checks its character and proves that it is a certified site before the end clients.

Index Terms - Phishing,Visual Cryptography,Image Captcha.

1. Introduction

Online exchanges are these days get to be extremely common and there are different assaults show behind this. In these types of different assaults, phishing is recognized as a major security threat and new creative thoughts are emerging with this in each second so preventive components ought to additionally be so effective [3]. Thus the security in these cases be high and ought not to be easily tractable with execution effortlessness Today, most applications are just as secure as their basic framework. Since the configuration and engineering of middleware has enhanced relentlessly, their location is a troublesome issue. Therefore, it is about difficult to make certain whether a machine that is associated with the web can be viewed as reliable and secure or not. Phishing tricks are likewise turning into an issue for internet keeping money and e-business users. The inquiry is the means by which to handle applications that oblige a high level of security.

Phishing is a type of online fraud that intends to steal sensitive data, for example, internet saving money passwords and Visa data from clients. Phishing tricks have been receiving far reaching press scope in light of the fact that such assaults have been heightening in number and advancement. One meaning of phishing is given as "it is a criminal movement utilizing social engineering procedures. Phishers endeavor to fraudulently acquire delicate data, for example, passwords and credit card subtle elements, by taking on the appearance of a dependable individual or business in an electronic correspondence". An alternate exhaustive meaning of phishing, expresses that it is "the act of sending an email to a client dishonestly guaranteeing to be an established honest to goodness venture into an endeavor to trick the user into surrendering private data that will be utilized for wholesale fraud". The behavior of fraud with this procured touchy data has additionally gotten to be simpler with the utilization of innovation and fraud can be depicted as "a wrongdoing in which the impostor acquires key bits of data such as social Security and driver's permit numbers and uses them for his or her own addition".

As the name describes, in this approach website cross verifies its own identity and proves that it is a genuine 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 idea of picture preparing and an enhanced visual cryptography is utilized. Picture preparing is a system of handling an information picture and to get the yield as either enhanced manifestation of the same picture and/or attributes of the data picture. In Visual Cryptography (VC) a picture is decayed into shares and keeping in mind the end goal to uncover the first picture suitable number of shares ought to be consolidated.

2. BACKGROUND

Phishing involves sending electronic mail or other appearance of trades to assembling of people asking for their individual information like Visa numbers and passwords. Aggressors makes the locales that almost look like sanction destinations and advances those destinations on Internet. Exactly when customer login through those destinations they are truly occupied to phisher's database where aggressor can get singular information of customer like watchword, record purposes of investment et cetera.

Regardless the aggressor makes a faked webpage in a web server. This site would have a striking similarity as the honest to goodness site.

2. Using a couple of gadgets they send stores of criticized e-sends to target customers for the purpose of genuine associations and affiliations, endeavoring to influence misused individuals to visit their destinations. 3. Exactly when the customer opens an email and clicked on the personification hyperlink the association will occupy them to a page asking the customer to enter the obliged information. 4. Once the customers incorporate their information, the phishers will get that information and can do anything they require with this information, including drawing out the money from the customers' record

3. Related Work

Online transactions are nowadays becoming very common, various different techniques are employed by an individual or group of people to perform fraudulent activities, which involves use of creating of fake sites which are similar to original site and is presented to user, to gain access to users personal information, middle man attacks are common where data is obtained while data traverse over network thus data security and phishing is an important area where work needs to be done. The below graph shows various phishing attacks being performed in last recent years.

I. CLASSIFICATION OF PHISHING ATTACK:

Phishing attacks can be classified in to following types according to the way attack is done

1. Deceptive phishing: In this type of phishing attacker broadcasts an email such as message regarding Need to verify account information, reenter user’s information because of system failure, undesirable

Account changes, new free service, and may other scams, with the hope that victim will enter their information and caught in to attackers trap.

2. Malware based phishing: This type of attack involves running malicious software on victims pc,malwares can be introduced as email attachment, or in downloadable file from website, or by exploiting security vulnerabilities.

3. Web Trojans: This kind of attacks pops up invisibly when users attempt to log in. They collect users. Information locally and transmits to the phisher.

4. System reconfiguration attack: In this type of attack users pc configuration is changed for malicious purpose to redirect users to the URL look alike, for example the Banks URL may be changed from www.gmail.com to www.gmai1.com, we can see here l is replaced by 1.

5. Man in middle phishing: In this type of phishing attacker puts themselves between the user and legal website, they record the user’s information and continue to the legal website so that user can not identify, user’s transactions are also not affected. Later the sell or use the user’s information when user is not active on the system.

6. Search engine phishing: In this type of phishing attacker creates very much attractive website with sound effects, so when users do normal search they find such kind of website and are fooled by giving up their information.

Many techniques came into existence to prevent phishing attacks they can be categorized as follows:

Identity Based Anti-Phishing Techniques

In this technique mutual identification is done where the user and visiting site checks each other identity while handshake. It is an anti-phishing technique which integrates partial credentials sharing and client filtering technique to prevent phishers from easily masquerading the online websites. As mutual authentication is followed, there would be no need for users to reenter their credentials. Therefore passwords are never exchanged between users and online entities except during the initial account setup process .

Advantage: It provide mutual authentication for server as well as client side. Using this techniques user does not to reveal his credential password in whole session except first time when the session is initialized [3].

Disadvantage: In identity based anti-phishing if a hacker gain access to the client computer and disable the browser plug-in then method will be compromise against phishing detection [3].

DNS based Black List and white list approach

DNS-based approach came to combat phishing [4] where this technique makes use of blacklist, where the Blacklist is a DNS based anti-phishing technique now most commonly used by the internet browsers. Almost all browsers use black list method to protect from phishing site which involves Google chrome, internet explorer, Firefox Mozilla, Netscape navigator etc., are important browsers which make use of blacklists technique to protect users when they are navigating through phishing sites.

Microsoft and AOL used Black List Approach [11][12] by integrating black list-based anti-phishing support into their browsers. It blocks users from entering any information while he/she is at a known phishing website. *Phishing Guard* [5] makes use of a white listing approach. The basic idea is that a website is identified by its IP address. The white list contains only trusted URLs. Whenever any website is visited, it is checked in the white list. If the URL is found in the white list, but the IP address is different than in list, then this URL is termed as phishing site. If the URL is similar but not the accurate one, then a warning/exception is provided at user end.

Advantage: it is most commonly used technique within web browser and easy to implement

Disadvantage: it is ineffective because there is always a chance of vulnerability during which users are susceptible to attacks.

Heuristic-based anti-phishing technique

Heuristic-based anti-phishing technique is to calculate whether a web page contains phishing heuristics characteristics. Some heuristics distinctiveness used by the Spoof Guard [6] toolbar include checking the machine name, checking the URL for available familiar spoofing techniques, and checking against earlier seen images. If we only use the Heuristic-based technique, the accuracy is not enough. Its pages are often similar with the legitimate sites. Therefore, some researchers proposed a correspondence assessment scheme to detect phishing sites.

Advantages: it provides an efficient checking mechanism where from hostname, URL, images are checked to detect phishing

Disadvantage: Heuristic-based anti-phishing technique, are having high probability of false alarm, and it's easy for an attacker to use technical means to avoid the heuristic characteristics detection.

Content based approach

CANTINA [7] is an content similarity based approach where data is stored. Firstly, it calculates the suspicious page's lexical signature using TF-IDF and then feed this lexical signature to a search engine. According to the suspicious page's sorted order in the search results we can determine whether it is a phishing site or not.

Advantage: In Content-based phishing detection is greater to detection using white and black lists because it does not require the maintenance of lists.

Disadvantage: The keyword extraction method (if- idf) presently used for content-based detection is insufficient and causes a high rate of false positives. This evaluation based strategy is tedious. It needs long time to figure a pair of pages, so utilizing this system to identify phishing sites on the customer Terminal is not suitable. Furthermore there will be low precision rate for this technique relies on upon numerous elements, such as the content, pictures, and comparability estimation strategy.

Attribute based anti-phishing techniques

Attribute-based anti-phishing strategy uses both reactive and proactive anti-phishing. This technique has been implemented in Phish Bouncer [8] tool. The Image Attribution technique does a comparison of images of accessing site and the sites already being registered with phish bouncer. The HTML Crosslink checks and looks at the responses coming from nonregistered sites and counts the number of links the page has to any of the registered sites. A high number of cross-links indicates that it is an phishing site[4]. In false info feeder[4] checker, false information is provided and if that information is accepted by the site ,then probably that link is phished one. It checks for suspicious certificates and validates site certificates presented during SSL handshake and extends the typical Usage by looking for Certification Authority (CA).As multiple checks are performed to authenticate the site this results in slow response time.

Advantage: As attribute based anti-phishing considers a lot of checks so it is able to detect more phished sites than other approaches. It can detect known as well as

Disadvantage: As multiple checks perform to authenticate site this could result in slow response time.

Genetic Algorithm Based Anti-Phishing Techniques

It is a methodology of location of phishing website pages utilizing hereditary calculation. Hereditary calculations can be utilized to develop straightforward principles for forestalling phishing assaults. These principles are utilized to separate ordinary site from phishing site. These phishing sites allude to occasions with likelihood of phishing assaults. The guidelines put away in the principle base are typically in the taking after structure [9]: if { condition } then { act } Case in point, a tenet can be characterized as: If the off chance that { The IP address of the URL in the got email Discovers any match in the Rule set} At that point {Phishing email } [9]

This rule can be explained as: if there exists an IP address of the URL in e-mail and it does not match the defined Rule Set for White List then the received mail is a phishing mail [9].

Advantage: It provides the feature of malicious status notification before the user reads the mail. It also provides malicious web link detection in addition of phishing detection.

Disadvantage: Single rule for phishing detection like in case of url is far from enough, so we need multiple rule set for only one type of url based phishing detection. Likewise for other parameter we need to write other rule this leads to more complex algorithm.

Later on Divya James and Mintu Philips proposed anti-phishing framework using visual cryptography where they referred work done by M. Naor and A. Shamir on visual cryptography and they used this technique to build anti-phishing framework. Work of Naor and Shamir is explained further.

Visual Cryptography

Until now the best technique to transfer data from one location to another securely over a network is cryptography where the sender encrypts the data and receiver decrypts the data such that only sender and receiver who are involved in communication can view the data. Cryptography involves complex mathematical calculations to perform these operations. Naor [11] and Shamir [11] introduced a new technique called visual cryptography (VCs) as a simple and secure way of sharing an image without any cryptographic computation. In visual cryptography an original image is divided in to n images to decrypt the n shares need to be stacked together to obtain the original image; even n-1 shares stacked together cannot reveal the actual image. Someone is enable to decrypt all n images are present otherwise n-1 images can not reveal an original image. When all images are overlaid then only original image can be produced.

Visual cryptography is an scheme in which visual information is encrypted by using encryption techniques where decryption can be done only by human visual framework, we can accomplish this by one of the accompanying plan

1. (2,2)Threshold VCS scheme-This is the simplest scheme that takes secret message and encrypts it in different shares and original message is revealed only when both the shares are available. no

additional information is needed here.

2. (2,n) Threshold VCS scheme: In this kind of scheme the secret image is encrypted in to n shares and

original image can be revealed when two (or more than two) shares are overlaid.

3. (n,n) Threshold VCS Scheme: in this kind of encryption scheme, it encrypts original image into n shares/sheets and original image can be revealed only when all the n shares are available simultaneously. The user will be prompted for n, the number of participants.

4. (k,n) Threshold VCS scheme: In this type of scheme the secret image is encrypted in to n shares ,at

time of formation of original image at least k group of shares must be present. The user will be prompted for k, the Threshold, and n, the number of participants.

Whereas more advance algorithms are proposed by using RGB pixel level displacement of images[1].

Using (2,2) Threshold VCS scheme Divya James, Mintu Philip, proposed “a novel anti-phishing framework using visual cryptography”[2] they divided the overall anti-phishing system into two modules registration and login phase, they are explained below.

4. PROPOSED METHOD

The proposed methodology has three phases. They are known as registration and authentication and data uploading/downloading phase.

A.Registration phase

User enters username and a key, server generates a random key and concatenate both the keys, and using the formed key a captcha image is stored on server. The image is divided into two shares one is kept with server and other with client, later on original image is send to client.

Registration process is depicted in Figure.1.

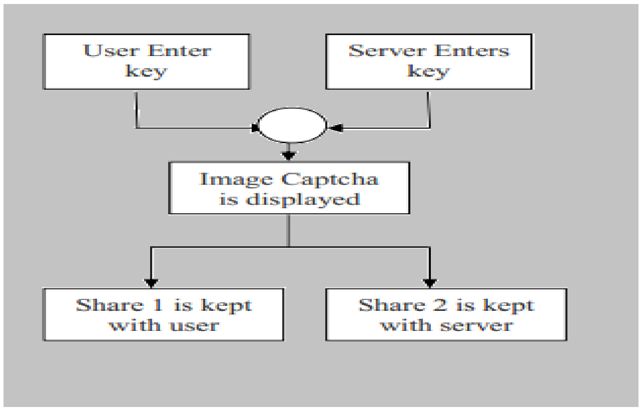

User Enter key
Server Enters key
Image Captcha is displayed
Share 1 is kept with user
Share 2 is kept with server

Figure 1.Registration process for the website

B. Login phase

client has to present his username and his share, when server share and client share is stacked together original image is obtained, then client/user can compare it with original image which he has If obtained image is not similar to image which user has it means the site is an phishing site. If original image is formed user can use this captcha image as a password to login. After user logins to actual site if user has to access data i.e. upload and download data he has.

Login process is depicted in Figure.2.

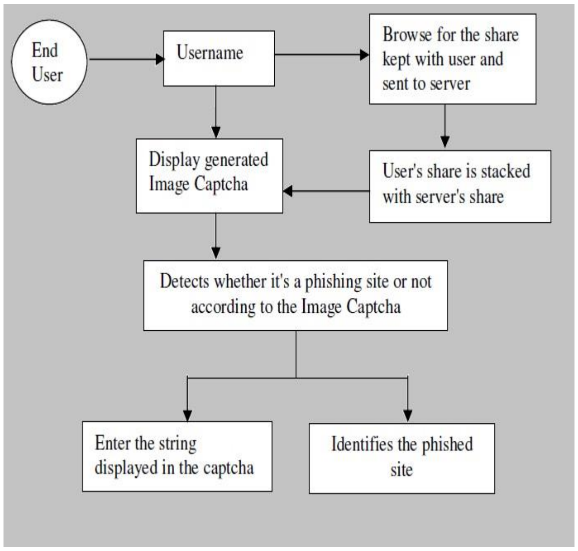

End User
Username
Browse for the share kept with user and sent to server
Display generated Image Captcha
User's share is stacked with server's share
Detects whether it's a phishing site or not according to the Image Captcha
Enter the string displayed in the captcha
Identifies the phished site

Figure 2. When user attempts to log in into site.

For generating the share on server and client I have used a new technique where image is ciphered using RGB pixel shuffling such that when inverse shuffling of RGB pixels are done original image is obtained.

The below algorithm depicts the overall process to generate shares

1. Import data from image captcha generated by server and create an image graphics object by interpreting each element in a matrix.

2. Get the size of r as [c, p]

Where c, p represents the width and height values of

The image

Remove the red component as a share r as an matrix

Remove the green component as a share g as an matrix,

Remove the blue component as a share b as an matrix.

Divide r, g, and b component matrix into two equal parts

Let first part be I, and second be j

I= (R1+G1+B1)

J= (R2+G2+B2)

For example

I'= Transpose I component

J'=Transpose g component

Add I' matrix pixels values into an image and form a new image which would be used as client share, also add J' matrix pixel values into image to form other new image which would be used as server share, when inverse operation is performed and RGB pixels are added back, when they would be stacked together performing inverse operation actual image would be obtained which could be used to detect phishing and would be used as an password for user.

5. MATHEMATICAL MODEL

For generating image shares mathematical model is as follows

Consider a set S_i whose values are taken into matrix

$S_i = \{R_i , G_i , B_i\}$

$$\begin{pmatrix} R_i & G_i & B_i \\ . & . & . \\ .R_n & G_n & B_n \end{pmatrix}$$

This is an set of image comprises of Red,Green,Blue pixels whose shares need to be created fur-ther ,the R,G,B pixel values of image are taken into matrix.

Further R_i component matrix is formed

S_1= {R_i} where {R_i}={R_0 ..R_n}

$$\begin{pmatrix} Ri \\ . \\ .Rn \end{pmatrix}$$

Further G_i component matrix is formed

S_2= {G_i} where {G_i}={G_0 ..G_n}

$$\begin{pmatrix} G_i \\ . \\ G_n \end{pmatrix}$$

Further B_i component matrix is formed

S_3= {B_i} where {B_i}={B_0 ..B_n}

$$\begin{pmatrix} B_i \\ . \\ B_n \end{pmatrix}$$

Further R_i ,G_i,B_i matrix are futher divided into two equal parts

R_a={$R_{0\ldots}R_{n/2}$}

R_b={$R_{n/2\ldots.}R_0$ }

G_a={$G_{0\ldots}G_{n/2}$}

G_b={$G_{n/2\ldots.}G_0$ }

B_a={$B_{0\ldots}B_{n/2}$}

B_b={$B_{n/2\ldots.}B_0$ }

Further transpose of these matrix are taken

R_{at}=transpose{R_a}

R_{bt}=transpose{R_b}

G_{at}=transposef{G_a}

G_{bt}=transpose{G_b}

B_{at}=transpose{B_a}

B_{bt}=transposef{B_b}

Further $R_{at,}G_{at,}B_{at}$ are combined together and add in image to form client share and R_{bt},G_{bt},$B_{bt\ are}$ combined to form image which is used as server share.

R_c={R_{at};G_{at};B_{at}}

R_c is the client share

R_s={R_{bt};G_{bt};B_{bt}}

R_s is the server share

Further client share and server share pixles are randomized further using java Random function, and seeds values are kept common.

When Inverse operation is perform random seed value are collect in array for both client and server and further below operations are performed.

R_a=transpose{R_{at}}

R_b=transpose{R_{bt}}

G_a=transpose{G_{at}}

G_b=transpose{G_{bt}}

B_a=transpose{B_{at}}

B_b=transpose{B_{bt}}

$R_{i\,=}${R_a U R_b}

G_i={G_a U G_b }

B_i ={B_a U B_b }

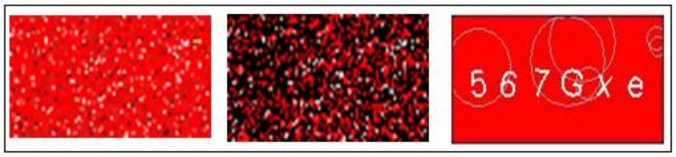
The original set of image is obtained S_i ={R_i ,$G_{i,}B_i$ }

Which is the original captcha image.

6. RESULTS

When user logins to phished site original image cannot be obtained, when client gives his share for login phished site also need to provide his share, which he doesn't have, even if he present an actual image captcha cannot be formed since other share is only known to server. And hence even if phished site presents an image share original image cannot be formed, a user can then compare the image formed with his original image he has.

Below figure depicts the process of formation of original image



(i) (ii) (iii)

Figure 3.Shows when Images are stacked how the original image is obtained.

In figure 3. (i) image shows client part (ii) image shows server part (iii) third image shows how after stacking both image captcha original image captcha is obtained else client image provided is not the exact image original image won't be obtained and user cannot login .

Orginal Image	Reconstructed image	Correlation cofficient
		0.978
		0.966
		0.968
		0.975

Table 1- Depicts correlation between two images after user authentication.

It is observed that both unique and reproduced picture captcha are connected with high level of relationship. The relationship coefficient of unique captcha and recreated captcha are indicated in Table 1.Also when two separate shares are stacked their relating connection co-Proficient is gotten as -0.0060.This demonstrates that there will be zero level of connection between unique and yield pictures for two separate shares.

7. CONCLUSIONS

The propose system guarantees protection from fraudulent sites and provides access to authenticated user, use of RGB pixel displacement concept further leverages security of client and server captcha shares being transferred over network. Since shares cannot be tampered or obtained in any case while traversing over network this system guarantees security from phishing sites and fraudulent activities.

Future Scope

The future work of this paper is the system can be used in banking and financial sites to avoid phishing attacks and protection of data.

REFERENCES

[1] Quist-Aphetsi K.ester, A New hybrid Asymmetric key-exchange Visual Cryptographic Algorithm tor Securing Digital images, 978-1 -4799-3067-8/13/$31.00 ©2013 IEEE

[2] Divya James, Mintu Philip, A Novel Anti Phishing framework based on Visual Cryptography, 978-1-4673-0446-7/$25.00©2012 IEEE

[3] Hicham Tout, William Hafner "Phishpin: An identity-based anti-phishing approach" in proceedings of international conference on computational science and engineering, Vancouver, BC, pages 347-352, 2009.

[4] 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.

[5] J. Kang and D. Lee, "Advanced white list approach for preventing access to phishing sites," in ICCIT '07: Proceedings of the 2007 International Conference on Convergence Ibenformation Technology. Washington, DC.USA: IEEE Computer Society, 2007, pp. 491-496.

[6] Nourian, A.; Ishtiaq, S.; Maheswaran, M.;" CASTLE: A social framework for collaborative anti- phishing databases", in Proceedings of IEEE- 5th International Conference on Collaborative Computing:Networking, Applications and Worksharing, 2009.

[7] Y. Zhang, 1. I. Hong, and L. F. Cranor, "Cantina: a content-based approach to detecting phishing web sites," in WWW '07: Proceedings of the 16th international conference on World Wide Web. New York,NY, USA: ACM, 2007, pp. 639-648.

[8] Michael Atighetchi, Partha Pal "Attribute-based prevention of phishing attacks" Eighth IEEE international symposium on network computing and application, 2009.

[9] V.Shreeram, M.Suban, P.Shanthi, K.Manjula "Anti-phishing detection of phishing attacks using genetic algorithm" in proceedings of Communication control and computing Technology(ICCCCT),IEEE international conference, Ramanathapuram , pages 447-450, 2010.

[10] M. Naor and A. Shamir, “Visual Cryptography,” Advances in Cryptology EUROCRYPT, 1994, Proceeding, LNCS vol.950, Springer-Verlag, 1995, pp. 1–12

[11] News.com, "Netscape readies antiphishing browser,"http://news.cnet.com, 2006.

[12] Microsoft, "Technology overview:microsoft windows internet explorer 7," Microsoft White Paper, 2006.

[13] Boldyreva, A., Degabnele, J. P., Paterson, K.Cr.& Stam, Jvl. (2012). Security of symmetric encryption in the presence of ciphertext fragmentation, in Advances m Cryptology-EUROCRYPT 2012 (pp. 682-6yy). Springer Berlin Heidelberg.

[14] Tianyang Li.; Fuye Han.; Shuai Ding and Zhen Chen.;"LARX: Large - scale Anti - phishing by Retrospective Data -Exploring Based on a Cloud Computing Platform", in Proceedings of IEEE-20th International Conference on Computer Communications and Networks, 2011.

[15] Qingxiang Feng.; Kuo-Kun Tseng.; Jeng-Shyang Pan.; Peng Cheng and Charles Chen.;"New Antiphishing Method with Two Types of Passwords in OpenID System", in Proceedings of IEEE Fifth International Conference on Genetic and Evolutionary Computing,2011

[16] Nirmal, K.; Ewards, S.E.V.; Geetha, K."Maximizing online security by providing a 3 factor authentication system to counter -attack 'Phishing'", in Proceedings of IEEE-International Conference on nEmerging Trends in Robotics and Communication Technologies, 2010.