ISSN No: - 2456- 2165

Improve Security for Fake clicks authentication by Unsupervised Captcha

Prashant Kumar Abhishek
M.Tech Scholar , Computer Science
Jaipur National University , Jaipur
prashant98015@gmail.com

Savita Shiwani
HOD, Computer Science
Jaipur National University , Jaipur
savitashiwani@gmail.com

Abstract:- In recent years, a considerable number of public services in the web have been trying to prevent abuse from automated programs by requiring from the users the resolution of a challenge in the format of a "Turing test" [12], popularly known as CAPTCHA, "Completely Automated Public Turing test to tell Computers and Humans Apart". In these services, the users are only able to start using the service after providing a successful answer to the test. The main advantage of leveraging such tests is that, theoretically, the tests are easily generated and answered by humans. All CAPTCHAs possess some sort of secret information which is initially only known by the challenger but not by the agent being challenged. In this Paper, we are working for improve the performance of the design Captcha so that security can be enhance. In the previous paper [18], they are working for the Captcha protocol. In which Captcha is design by the graphical representation. User has to see the Captcha and type the same text into the given box. We are using unsupervised Captcha which divide the given input image in Auxillary visual word, Bow model, Affinity matrix, image annotation. By this process the execution time will get reduce and security of the Captcha will get improve.

Keyword:- *CARP, E-mail attacks, Online guessing attacks, Relay attacks, Shoulder surfing attack.*

I. INTRODUCTION

In recent years, a considerable number of public services in the web have been trying to prevent abuse from automated programs by requiring from the users the resolution of a challenge in the format of a "Turing test" [12], popularly known as CAPTCHA, "Completely Automated Public Turing test to tell Computers and Humans Apart". In these services, the users are only able to start using the service after providing a successful answer to the test. The main advantage of leveraging such tests is that, theoretically, the tests are easily generated and answered by humans. All CAPTCHAs possess some sort of secret information which is initially only known by the challenger but not by the agent being challenged. The scheme necessary to implement our solution is based on the authentication of users after resolution of a CAPTCHA, aiming to distinguish human users from computer bots. Once the CAPTCHA is successfully solved, the user receives a ticket in the form of a cookie [13]. For security reasons, the cookie has an expiration date and time, after which the user needs to answer a new test. The ticket and all its related information could be stored locally as a user session variable or could be implemented differently, in a central authority that would keep record of valid active tickets.

A. Architecture

In a traditional scheme, when a user clicks on an ad located in the site of a publisher, the corresponding ad is obtained from the advertisesr and the transaction is recorded by the advertising network. After this, the advertising network will charge the advertiser and pay the publisher. In the model here proposed, there are some additional tasks to be performed: when the user clicks on an ad (for example, in an advertising network banner located on the site of a publisher), it is prompted by the advertising network with a new web page containing a clickable CAPTCHA that needs to be solved. If the challenge is not answered correctly, a new challenge will be proposed, and the ad will not be exhibited. This process will be repeated until the challenge is solved correctly. Once this happens, a ticket certifying that the user is human is embedded by the advertising network in the user browser, the advertising network records the action and the ad is finally displayed to the user. Whenever a previously authenticated user clicks in any ad, the ticket will be automatically sent to the advertising network, triggering a validation process in its servers. After the validation and confirmation that the ticket sent is valid, the ad is displayed. To mitigate the risk involving a situation in which a user authenticates once and then executes a script that will run overnight from their workstation, the tickets are only valid during a short period of time, after which it is necessary to perform a new authentication. It is important to observe that employing the use of cookies or tickets in the current filtering methods used for click-fraud detection is not indicated. This is because the filtering process is an exclusive one: if one cookie was used to mark and exclude a determined type of malicious user, fraudsters could simply remove the cookies from their web browsers. This is why we consider this methodology "separative", since it accepts only valid clicks, it may very well make use of cookies. The tickets mark good, trustable users. This is the same reason why the mechanism here proposed is prevention-based instead of detection-based.

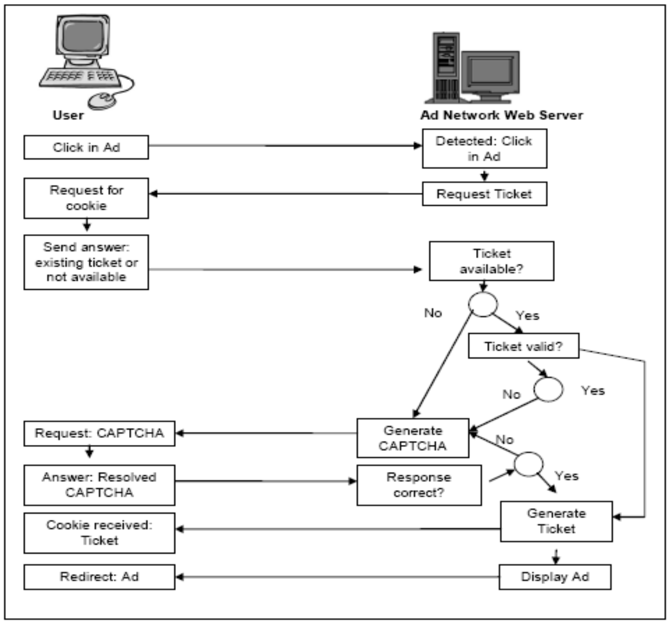


Fig 1. Communication Flow for Ticket Authentication

II. RELATED WORK

In general, there is extensive literature on captcha and graphical system to avoid machine learning attacks. This section reviews about the some related work in order to explore the strengths and weakness of existing methods. P.C.Van Oorschot, A. Salehi- Abari, and J. Thorpe [1] this paper proposes a purely automated attack on pass points-style graphical passwords. Which are easier to arrange than human-seeded attacks and more scalable to systems that use multiple of image. It requires serious consideration when deploying basic Pass Points-style graphical passwords and possible of trail to gain password. M. Alsaleh, M. Mannan, and P.C.Van Oorschot [2] this paper proposes a Revisiting defences against large-scale online password guessing attacks Easy-to-deploy approach to identify automated malicious login attempts with reasonable cost of inconvenience to users. The third party human attack employs hired human to solve challenges so that the CAPTCHA systems will no longer be effective. It produces online guessing attack. G. Moy, N. Jones, C. Harkless, and R. Potter [3] this paper proposes Distortion estimation techniques in solving visual captchas. Estimation technique is used to measure the attacks. This captcha test cannot pass the machine. A direct distortion an estimation algorithm that correctly an identified a four letters in a challenge image 78% only. It vulnerable to brute force attacks. P.C.Van Oorschot, Julie Thorpe [4] this paper proposes an on predictive models and user-drawn graphical passwords. To better understand the size of these classes, how weak the password subspace. Motivate us to define a set of password complexity factors which define a set of classes. Thus, it is possible that if the system had protected information that was perceived to be sensitive. Some of these users might have created passwords they perceived to be more complex. B.B. Zhu [5] this paper proposes an Attack and design of image recognition captchas an unlimited number of types of objects can be used in Cortcha. No need to manually label any image and strength of Learn ability and efficiency. An infinite number of object types are used to generate Cortcha challenges. Cortcha does not require the images in its image database to be labeled.

III. CAPTCHA PROTOCOL

A primary job of Captcha as a graphical password can provide security. For example the RSA algorithm is developed based on factorization problem and elliptic curve, DSA-digital signature algorithm, Elgamal algorithm, Diffie Hellmen algorithm is developed based on the problem of the Discrete logarithm problem. It is based on the AI Problem, they can also create CARP technology from the problem of captcha. It is used to detect the user where the computer used by human or machine. We develop the relations of CARP –Captcha as graphical password. CARP is click-based password. Where continues click on the particular image used to generate password. Compare with other password, CARP password provide more security. CARP technology can provide online and e-mail security by using the Text captcha, Click Animal, Animal Grid.

Every login of CARP a new image is generated. Text captcha and an image captcha are used in CARP schemes. It looks like same as text password of sequence of characters. The entered value can be change by clicking on the image of characters. CARP provides protection and restrict the online dictionary attacks on the password. Now days various online service and attacks arises by using CARP to provide security. This should be top of cyber security risks because of the threats is widespread. The subtle problem of online dictionary attacks is might appear. CARP is also used to provide the security against relay attacks. The CARP images are answered by human and machine cannot to do. In dual view technologies are used to against shoulder-surfing attacks and CARP also provide robust. The CARP image is difficulty for machine. The only required is solving the CARP image in every login. CARP is a collection of Captcha and graphical system. First we are known about captcha and graphical password. CAPTCHA is an acronym for Completely Automated Public Turing Test to tell Computers and Human Apart. Captcha is used to find the computer used by the user or machine. CAPTCHAs also hand out as a standard job for artificial intelligence technologies. CAPTCHA can be second-hand to answer a hard unsolved AI problem. The problems are unsolved means the system used by automaton. If an AI were skilled of correctly realization the task without exploiting flaws in an exacting CAPTCHA blueprint, then it would have solved the difficulty of increasing an AI that is talented of compound object credit in scenes.

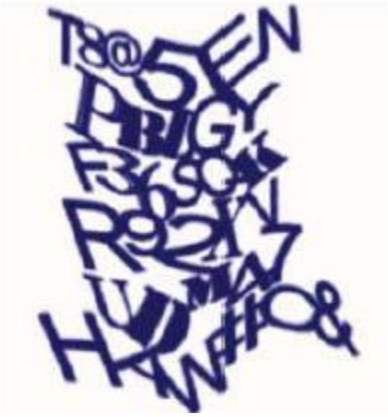


Fig 2:- Demonstrates Click Text.

IV. PROPOSED METHODOLOGY

For improve the security and reduce the elapse time , we are introducing the unsupervised based Captcha which complete the process in the form of Auxillary visual word , Bow model , Affinity matrix , image annotation .

- *Step1:-* In the unsupervised Captcha we are maintaining the database of the different types of images.
- *Step 2:-* It will do the process of Auxiliary Visual words generation for matching points.
- *Step 3:-* Then we are adopting the Bow method to detect the foreground and background.
- *Step 4:-* Then we are adopting the Affine matrix to convert it in to pattern.
- *Step 5 :-* Now we are applying the Annotation to get the feature vectors to get the relevant image from the database.
- *Step 6 :-* When it will find the relevant image it will done the process of verification
- *Step 7 :-* This method is universal and purely depend up on the specific database used for very high security.

A. Bow Model

In computer vision, the bag-of-words model (BOW model) can be applied to image classification, by treating image features as words. In document classification, a bag of words is a sparse vector of occurrence counts of words; that is, a sparse histogram over the vocabulary. In computer vision, a bag of visual words is a vector of occurrence counts of a vocabulary of local image features.

B. Image Representation Based On The Bow Model

To represent an image using BOW model, an image can be treated as a document. Similarly, "words" in images need to be defined too. To achieve this, it usually includes following three steps: feature detection, feature description, and codebook generation. A definition of the BOW model can be the "histogram representation based on independent features". Content based image indexing and retrieval (CBIR) appears to be the early adopter of this image representation technique.

C. Feature representation

After feature detection, each image is abstracted by several local patches. Feature representation methods deal with how to represent the patches as numerical vectors. These vectors are called feature descriptors. A good descriptor should have the ability to handle intensity, rotation, scale and affine variations to some extent. One of the most famous descriptors is Scale invariant_feature transform (SIFT). SIFT converts each patch to 128-dimensional vector. After this step, each image is a collection of vectors of the same dimension (128 for SIFT), where the order of different vectors is of no importance.

D. Codebook generation

The final step for the BOW model is to convert vector-represented patches to "code words" (analogous to words in text documents), which also produces a "codebook" (analogy to a word dictionary). A codeword can be considered as a representative of several similar patches. One simple method is performing k-means clustering over all the vectors. Code words are then defined as the centers of the learned clusters. The number of the clusters is the codebook size (analogous to the size of the word dictionary).

Thus, each patch in an image is mapped to a certain codeword through the clustering process and the image can be represented by the histogram of the code words.

E. Image Annotation

Automatic image annotation (also known as automatic image tagging or linguistic indexing) is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. This application of computer vision techniques is used in image retrieval systems to organize and locate images of interest from a database.This method can be regarded as a type of multi-class image classification with a very large number of classes - as large as the vocabulary size. Typically, image analysis in the form of extracted feature vectors and the training annotation words are used by machine learning techniques to attempt to automatically apply annotations to new images. The first methods learned the correlations between image features and training annotations, then techniques were developed using machine translation to try to translate the textual vocabulary with the 'visual vocabulary', or clustered regions known as blobs. Work following these efforts has included classification approaches, relevance models and so on.

The advantages of automatic image annotation versus content-based image retrieval (CBIR) are that queries can be more naturally specified by the user. CBIR generally (at present) requires users to search by image concepts such as color and texture, or finding example queries. Certain image features in example images may override the concept that the user is really focusing on. The traditional methods of image retrieval such as those used by libraries have relied on manually annotated images, which is expensive and time-consuming, especially given the large and constantly growing image databases in existence.

V. RESULTS

Figure 3 is showing the Captcha protocol and Unsupervised Captcha for detect the authentic user. Figure 3 is showing the design GUI for the Captcha protocol and Unsupervised Captcha detection .

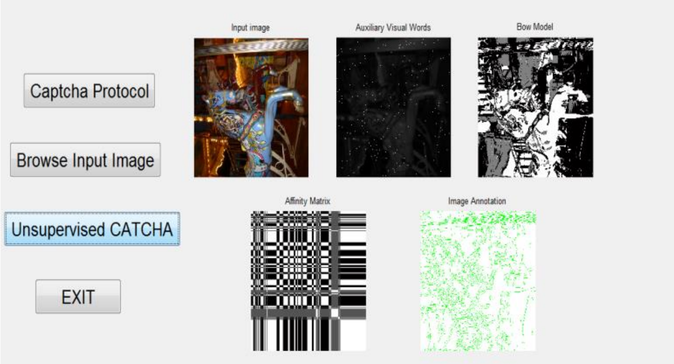
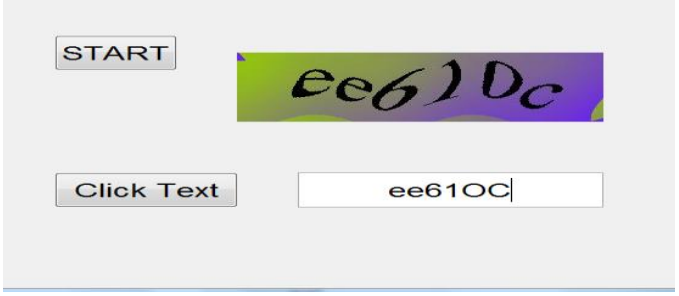


Fig 3. GUI for Captcha Protocol

As we can see from the figure 4 the Captcha protocol is working for detection the text . In the figure 3 , as we click at the captcha protocol then the new window will open as the figure 4. In this we will get the image of the text . If we give the same and correct text in the text window then it will generate the corresponding elements and match that . If both will get same then authentication process will successfully complete .



ee61OC

Fig 4. Captcha Protocol

For authentication from the unsupervised Captcha , the given image will be break in three samples that's are Auxillary visual word , Bow model , Affinity matrix , image annotation . Figure 5 is showing the comparison for the simulation time in between existing Captcha Protocol and proposed Unsupervised Captcha Protocol. As we can see from the figure 5, the execution time for the proposed unsupervised algorithm is low as compare to the Captcha Protocol.

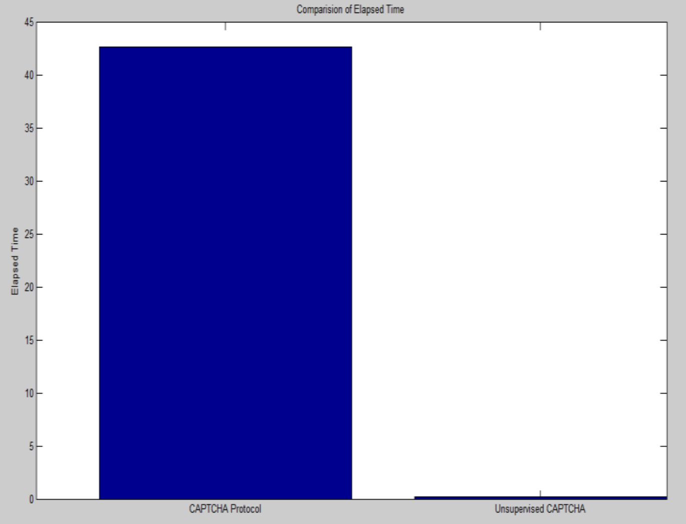


Fig 5. Elapsed time for comparison in CAPTCHA Protocol and Unsupervised Captcha

VI. CONCLUSION & FUTURE SCOPE

As we can see from the results session, the execution time is low for the proposed unsupervised algorithm as compare to the Captcha protocol. That means we can say that for check the authentic click unsupervised algorithm is best as compare to captcha protocol. We presented a click fraud prevention method, an innovation if compared to the great majority of the current click-fraud combat methods, which treat the fraud after its occurrence through the filtering and detection of fraudulent clicks. Unsupervised Captcha method, the approach here presented proposes the use of differentiation tests between humans and computers, through the use of clickable Class II CAPTCHAs. The answer to these tests will work as a validation certificate of the clicks which, after considered valid, will be accounted for further charge. In an ideal world, the method here proposed could surpass the current detection mechanisms; however it is specially indicated to be utilized in a complementary fashion. In the Future, we can work at the limitations of the project .As we can see that in the limitations, the given process is offline still now. We can progress it for the online. For authentication images are limited. So we can increase the images so that some different type of authentication can be performing. In the project, we are using just some limited images for authentications. These images can be increase in the future.

REFERENCES

[1] Bin B. Zhu, Jeff Yan, Guanbo Bao, Maowei Yang, and Ning Xu_, ―Captcha as Graphical Passwords— A New Security Primitive Based on Hard AI Problems‖ IEEE *Trans. Inf. Forensics Security*, Vol. 9, No. 6, June 2014.

[2] P. C. van Oorschot, A. Salehi-Abari, and J. Thorpe, ―Purely automated attacks on passpoints-style graphical passwords,‖ *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 393–405, Sep. 2010.

[3] M. Alsaleh, M. Mannan, and P. C. van Oorschot, ―Revisiting defenses against large-scale online password guessing attacks,‖ *IEEE Trans. Dependable Secure Comput.* vol. 9, no. 1, pp. 128–141, Jan./Feb. 2012.

[4] P. C. van Oorschot and J. Thorpe, ―On predictive models and user-drawn graphical passwords,‖ *ACM Trans. Inf. Syst. Security*, vol. 10, no. 4, pp. 1–33, 2008.

[5] G. Moy, N. Jones, C. Harkless, and R. Potter, ―Distortion estimation techniques in solving visual CAPTCHAs,‖ in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jul. 2004, pp. 23–28.

[6] B. B. Zhu et al., ―Attacks and design of image recognition CAPTCHAs,‖ in Proc. ACM CCS, 2010, pp. 187–200.

[7] P. C. van Oorschot and J. Thorpe, ―Exploiting predictability in click-based graphical passwords,‖ *J. Comput. Security*, vol. 19, no. 4

[8] J. Elson, J. R. Douceur, J. Howell, and J. Saul, ―Asirra: A CAPTCHA that exploits interest-aligned manual image categorization,‖ in *Proc. ACM CCS*, 2007, pp. 366–374.

[9] B. Pinkas and T. Sander, ―Securing passwords against dictionary attacks,‖ in Proc. ACM CCS, 2002, pp. 161–170.

[10] G. Mori and J. Malik, ―Recognizing objects in adversarial clutter,‖ in Proc. IEEE Comput. Society Conf. Comput. Vis. Pattern Recognit., Jun. 2003, pp. 134–141.

[11] M. Szydlowski, C. Kruegel, and E. Kirda, ―Secure input for web applications,‖ in Proc. ACSAC, 2007, pp. 375–384.

[12] G. Wolberg, ―2-pass mesh warping,‖ in Digital Image Warping. Hoboken, NJ, USA: Wiley, 1990.

[13] HP TippingPoint DVLabs, New York, NY, USA. (2011). The Mid-Year Top Cyber Security Risks Report [Online]. Available:http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA3-7045ENW.pdf.

[14] S. Kim, X. Cao, Hs. Zhang, and D. Tan, ―Enabling concurrent dual views on common LCD screens,‖ in Proc. ACM Annu. Conf. Human Factors Comput. Syst., 2012.

[15] S. Li, S. A. H. Shah, M. A. U. Khan, S. A. Khayam, A.-R. Sadeghi, and R. Schmitz, ―Breaking e-banking CAPTCHAs,‖ in Proc. ACSAC, 2010, pp. 1–10.

[16] H. Gao, X. Liu, S. Wang, and R. Dai, ―A new graphical password scheme against spyware by using CAPTCHA,‖ in Proc. Symp. Usable Privacy Security, 2009, pp. 760–767.

[17] L. Wang, X. Chang, Z. Ren, H. Gao, X. Liu, and U. Aickelin, ―Against spyware using CAPTCHA in graphical password scheme,‖ in Proc. IEEE Int. Conf. Adv. Inf. Netw. Appl., Jun. 2010.

[18]. Revathi.M, Dhanalakshmi.S," *A New Captcha Protocol For Avoiding Machine Learning Attacks*", International Journal For Trends In Engineering & Technology Volume 3 Issue 3 – March 2015.

[19]. Monika Chilluru1, B. Ravindra Naick2, P. Nirupama ," Captcha based Password Authentication – A New Security Scheme", International Journal of Computer Science and Information Technologies, Vol. 6 (4) , 2015.

[20]. Carlos Javier Hernández-Castro, David F. Barrero, and María D. R-Moreno," AMachine Learning Attack against the Civil Rights CAPTCHA", Springer International Publishing Switzerland 2014 .

[21]. Bin B. Zhu1 and Jeff Yan2," *Towards New Security Primitives Based on Hard AI Problems*", Security Protocols 2013.

[22]. Goran Martinovic, Zdravko Krpic," *Advanced Character Collage CAPTCHA*", Advanced Character Collage CAPTCHA, 2012 .

[23]. Guido Schryena, Gerit Wagnera, Alexander Schlegel," *Development of two novel face-recognition CAPTCHAs: a security and usability study*",2016 .

[24]. Colin Hong, Bokil Lopez-Pineda, Karthik Rajendran, Adria Recasens," *Breaking Microsoft's CAPTCHA*".

[25]. Jeff Yan, Ahmad Salah El Ahmad," *Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms*".

[26]. Vikas K. Kolekar, Milindkumar. B. Vaidya," *A Review of Captcha and Graphical Passwords to Enhance Security and Usability to Next Level*", International Journal of Science and Research (IJSR), Volume 4 Issue 5, May 2015.

[27]. Ragavi.V, Dr.G.Geetha," *CAPTCHA Celebrating its Quattuor decennial – A Complete Reference*", IJCSI

International Journal of Computer Science Issues, Vol. 8, Issue 6, No 2, November 2011.

[28]. Takashi Tsuchiya, Masahiro Fujita, Kenta Takahashi," *Secure Communication Protocol Between a Human and a Bank Server for Preventing Man-in-the-Browser Attacks*", International Publishing Switzerland 2016.

[29]. Shujun Li, S. Amier Haider Shah, M. Asad Usman Khan, Syed Ali Khayam," *Breaking e-Banking CAPTCHAs*", 2010.

[30]. P. Kalaivizhi V. Udhayakumar," *A New Security Primitive Based on Carp Using Hard Ai Problems*", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 4, 2015.

[31]. Vijay Dhaka, Geeta Gandhi," *Developing a CAPTCHA Utilizing Cognitive Ability of Human through PHP*", Special Conference Issue: National Conference on Cloud Computing & Big Data.