CAPTCHA: A Security Review

Sudarshan Soni
Dept. of Computer Science and Engineering
Shri Shankaracharya Technical Campus
Junwani, Bhilai(C.G.)

Padma Bonde, PhD
Dept. of Computer Science and Engineering
Shri Shankaracharya Technical Campus
Junwani, Bhilai(C.G.)

ABSTRACT
Completely Automated Public Turing test to tell Computers and Human Apart(CAPTCHA) is an automated test that humans can pass, but current computer programs could not pass. The main purpose of CAPTCHA is to block automated scripts that are posted spam content everywhere they can. In this paper, various types of CAPTCHA techniques has been discussed. There is a constant require to enhance current CAPTCHAs and to develop new CAPTCHAs in order to provide security against newly generated programs which can create thousands of e-mail accounts used for stuff online polls with ballots, malicious purposes, and develop worms and viruses contained in emails.

Keywords
CAPTCHA, Graphical Password Authentication, Hard AI Problem, Attack.

1. INTRODUCTION
There are numerous sites which utilizes CAPTCHAs to endeavor to square computerized communications with their site. These endeavors might be extremely urgent to the accomplishment of these destinations in various ways. Here the illustration, Gmail show signs of improvement by its administrations by blocking access to the computerized spammers, eBay site show signs of improvement by its commercial center by blocking bots from flooding the site with tricks, and furthermore Facebook limits formation of fake profiles utilized by spam legit clients or cheat at diversions. The most regularly utilized CAPTCHA plans is to utilize blends of bended characters and muddling systems that people can perceive yet that might be confounded for mechanized scripts.

The term CAPTCHA depicting a test that can segregate people from PCs. Content CAPTCHAs are completely utilized as a part of true applications. In a content CAPTCHA, characters are purposefully twisted and associated with secure from acknowledgment by bots. A considerable lot of the proposed or actualized content CAPTCHA's have been broken. It is conceivable to expand the security of a current content CAPTCHA by reliably including clamor and mutilation, and also orchestrating characters all the more firmly. In any case, these measures will likewise make the characters harder for people to recognize, bringing about a higher blunder.

CAPTCHAs are in some cases otherwise called Switch Turing tests: since they are proposed to enable a PC to represent if a remote customer is human or not. Though, of their significance, that to a great degree utilized over far reaching, and a creating number of research studies there is as of now no predictable philosophy for outlining or assessing CAPTCHAs. Truly, as client substantiate by intensive review diverse mainstream sites still depend on plans that are defenseless to robotized assaults.

Examination of the subsequent information shows that CAPTCHAs are frequently entangled for people, with sound CAPTCHAs being especially more risky. Statistic patterns appears, for instance, that intrusive speakers of English are slower when all is said in done and less right on English-driven CAPTCHA plans. There are a few impediments to the mutilation and furthermore for the clamor that people can endure in an issue of content CAPTCHA. Ease of use is dependably an essential test in planning a CAPTCHA. With advances of division and Optical Character Acknowledgment (OCR) innovations, there is an ability hole amongst people and bots in perceiving bended and associated characters turns out to be progressively littler. This kind of pattern would likely render content CAPTCHAs in the end less successful.

Seeking elective procedures in planning CAPTCHAs to supplant content CAPTCHAs has turned out to be more imperative. A noteworthy exertion has been taken to creating CAPTCHAs in light of picture or question acknowledgment. Pictures are rich in data, natural to people, and of an immense variety. All the more significantly Pictures is by all accounts a superior channel than characters for outlining CAPTCHAs. To accomplish human learning and furthermore to make data more open to the world, different ventures are at present digitizing physical books that were composed before the PC age. The book pages have being photographically filtered, and after that, to make them analyze, changed into content utilizing Optical Character Acknowledgment(OCR).

2. LITERATURE SURVEY
[1] had discussed both CAPTCHA and password in a user authentication protocol, as a *Captcha-based Password Authentication (CbPA) protocol*, to counter online dictionary attacks.

[2] discussed two distortion estimation techniques that solved EZ-Gimpy and 4-letter Gimpy-r CAPTCHAs through object recognition with a high degree of success. Doing so they achieved a success rate of 99%. In the case of Gimpy-r a success rate of 78% was achieved by deploying a direct distortion estimation algorithm that was able to correctly identify the four lettered Gimpy-r CAPTCHA.

[3] have done a thorough study of visual CAPTCHAs which are available at captchaservice.org. It is a website that provides services for CAPTCHA generation publically which had sophisticated distortions and were meant to be resistant to OCR attacks.

[4] have presented a character segmentation technique to attack a number of text CAPTCHAs, including those designed and deployed by Yahoo, Microsoft and Google. Specifically they have targeted the Microsoft CAPTCHA which had been deployed since 2002 at a number of their own internet services including Windows Live, MSN and Hotmail. They implemented the attack in Java on an ordinary desktop computer (with a 1.86 GHz Intel Core 2 CPU and 2 GB

RAM). The average speed to completely segment a challenge was recorded to be slightly more than 80 milliseconds. They estimated that the given Microsoft CAPTCHA could be broken instantly by a malevolent bot with an estimated (segmentation and then recognition) success rate of 60% or more. However, the goal at the time of designing for the Microsoft CAPTCHA was that automated attacks should not achieve a success rate of more than 1 in 10000 attempts (that is, 0.001%). For the first time, the vulnerability of a segmentation resistant text-based CAPTCHA was exposed to different and simple attacks.

[5] proposed CAPTCHA Based on Image Orientation; The presentation of a new CAPTCHA which is based on identifying an image's upright orientation. This task requires analysis of the often complex contents of an image, a task which human usually perform well and machines generally do not.

A large number of graphical password schemes have been proposed. They can be classified into three categories according to the task involved in memorizing and entering passwords: recognition, recall, and cued recall. Each type will be briefly described here. More can be found in a recent review of graphical passwords [6].

A typical scheme is Pass faces [7] wherein a user selects a portfolio of faces from a database in creating a password. During authentication, a panel of candidate faces is presented for the user to select the face belonging to her portfolio. This process is repeated several rounds, each round with a different panel. A successful login requires correct selection in each round. The set of images in a panel remains the same between logins, but their locations are permuted.

[8] proposed Uncover impact factors of text-based CAPTCHA identification Presently, CAPTCHA is an important mechanism to gain access to the required system. However, there are some difficulties for users in typing CAPTCHA although they are authorized persons. The Text-based CAPTCHA is the most popular mechanism amongst all the CAPTCHA techniques, the difficulties of this text-based are studied and drawn out. The results of this study have shown that the presented character(s), genders of users, and their educational background are some of the important factors determining the correctness of CAPTCHA typing by its users. Therefore, generating a Text-based CAPTCHA must use the appropriate character, that also combining with the educational background and genders of the users.

[9] proposed an innovative image-based CAPTCHA for distinguishing human and computer by embedding versatile characters in the images and in method who they proposed it makes the characters invisible by automated image analysis technologies like scale-invariant feature transform while human can easily distinguish the location of the embedded characters .

[10] proposed a new paradigm has achieved just a limited success as compared with the cryptographic primitives based on hard math problems and their wide applications. Is it possible to create any new security primitive based on hard AI problems? This is a challenging and interesting open problem. Here introduced a new security primitive based on hard AI problems, namely, a novel family of graphical password systems integrating CAPTCHA technology, which we call *CaRP (Captcha as gRaphical Passwords)*.

The ROOT based CAPTCHA proposed in [11] consisted a large alphanumeric character, called *Base Character* with small variable sized characters placed over it at random positions. The number of characters placed over the *Base Character* varies within a range of 3 to 8. The reading pattern of the CAPTCHAs is specified in two ways: Navigation Instruction Scheme (NIS) and Navigation Line Scheme(NLS). By changing the instruction in NIS or guiding line in NLS, the same CAPTCHA can be used repeatedly with different answers each time, hence avoiding replay attacks to an extent. The resistivity of ROOT based CAPTCHA towards segmentation and letter recognition

[12] proposed a novel anti-bot mechanism called Necklace CAPTCHA for securing OSNs against the Social Bots. Social bots that exactly mimic the social behaviors such as auto-Likes, auto-photos/videos sharing, auto-sending friend requests, or auto-joining to strange groups. The ability of these dangerous bots is to perform those malicious activities reflects a big vulnerability in the authentication system of online social networks

General codes are directly crawled from the site ,the purpose of the [13] is designed to prevent the machine automatically identify, the code image has a lot of interference, such as noise, interference lines, twisted, distorted and so forth. Background and technical of validation code recognition is mainly based on image processing and pattern recognition techniques. That is prior to preprocess the image before the code image recognition, wherein the image pre-processing technology includes a gray-scale image, binaryzation, denoising, tilt correction, character segmentation and normalization,etc;.

3. METHODOLOGY USED

The CbPA-protocol in [1] requires solving a CAPTCHA challenge after inputting a valid pair of user ID and password unless a valid browser cookie is received. For an invalid pair of user ID and password, the user has a certain probability to solve a CAPTCHA challenge before being denied access.

[2] have performed different attacks on the given visual CAPTCHAs. For EZ-Gimpy they developed a correlation algorithm which uses “core” and “minipatch” framework and their variations to determine distortions and subsequently find which distorted template image best correlated to the challenge image.

Instead of using complex machine learning algorithms or computer vision techniques, [3] implemented simple pattern recognition algorithms along with serious design faults found in every scheme to break the CAPTCHAs with close to 100% success rate.

The existing CAPTCHA system in [4] was thought to be segmentation-resistant. However, their simple attack which involved 7 consecutive steps, namely – binarization, fixing broken characters, vertical segmentation, color filling segmentation, thick arc removal, locating connected characters and segment connected characters achieved a segmentation success rate of 92% against the scheme, that is out of 100 challenges 92 where segmented correctly.

In [5] given a large repository of images, such as those from a web search result, user can use a suite of automated orientation detectors to prune those images that can be automatically set upright easily. Then apply a social feedback mechanism to verify that the remaining images have a human-recognizable upright orientation. This CAPTCHA lends itself to rapid implementation and has an almost limitless supply of

images. [9] embed versatile characters in the images and in method who they proposed it makes the characters invisible by automated image analysis technologies. Their designed mechanism was capable to elude such generous of attacks. For in experiments, 15 users were invited to test the system and the success rate is 86%. If wrong operations like clicking out of text boxes were excluded, the success rate reached 95%. Compare the average logging time with reCAPTCHA and hello CPTCHA, the proposed method is faster than these two methods by 32 seconds and 115 seconds, respectively. In this paper they proposed random number generation module is used to select image and verification characters for conducting man machine distinguishing. Then color, size, style, and angle variation are applied to the verification characters. Finally, the verification image is generated.

The Text-based CAPTCHA used in [8] is the most popular mechanism amongst all the CAPTCHA techniques, the difficulties of this text-based are studied and drawn out. The results of this study have shown that the presented character(s), genders of users, and their educational background are some of the important factors determining the correctness of CAPTCHA typing by its users. Therefore, generating a Text-based CAPTCHA must use the appropriate character, that also combining with the educational background and genders of the users. The ROOT based CAPTCHA in [11] has characters placed at random positions with overlapping. Also, the characters have pixel discontinuities added to them without reducing usability. So, to an extent, the ROOT based CAPTCHA can resist segmentation and pixel attacks.

In [10] CaRP is click-based graphical passwords, where a sequence of clicks on an image is used to derive a password. Unlike other click-based graphical passwords, images used in CaRP are CAPTCHA challenges, and a new CaRP image is generated for every login attempt. Number of graphical password scheme used in [6] in also a powerful mechanism where recognition, recall and cued recall whereas in [7] portfolio of faces used in pass faces in which, user select a portfolio of faces from a database in creating a password.

The major functionality of the Necklace CAPTCHA in [12] depends on two approaches:

1) The Necklace Graph which used for visualizing and generating the CAPTCHA's Puzzles (or tests).

2) Vidoop CAPTCHA which uses images of objects (e.g. animals, birds, devices, etc) to add a noisy environment and to increase the semantic of our proposed CAPTCHA's challenge against the automated Scripts.

The basic objective of our proposed CAPTCHA mechanism is to provide a high level of usability and satisfaction to users while improving the robustness and security against the social bots in Social Networks. In the following, we provide a high level overview of the design and the methodology of our proposed Necklace CAPTCHA mechanism.

Relative shape context descriptors compared with the shape context descriptors in [13] has some obvious advantages: (1).not need the special selected polar coordinates of each point in the positive direction to reach the full rotation invariance, hence, has translational and rotational invariance; (2).not necessary to calculate items logρ,for all the points just only to be considered its positive direction of the polar coordinate polar angle, so possess the scale invariance; (3).not necessary to calculate the average distance from set point to all point to realize the regulation of distance entry, therefore, it has special point and point noise robustness. Pattern matching on [13] was based on a measure method of measuring similar to the shape. There are many Shape matching methods according to the different classification methods, for example, according to its ability of dealing with the transformation point to divided as follows: firstly, by searching invariant of various changes to deal with shape charge, these invariants are: Imitation shot invariant, similar invariants, perspective invariants. Secondly, by finding local features between the target and the samples to obtain the corresponding minimum match tolerance to handle more complex deformation

4. LIMITATIONS ON EXISTING SYSTEMS

The text based CAPTCHA use in [8] in most of system, the chances of failure is 92%, since computer automated spammer can access it easily. Whereas image used in [5] and [9] are sometimes even unable to recognize by human itself. As different users may take different time based on their abilities, so time bound can creates major problem in [11].

Graphical password of [6] and [10] have limited success as compared with the cryptographic primitives based on hard math problems and their wide applications.

In [2], [4] and [3] the success rate is also low due to the recognition in text CAPTCHA of distorted portion of text can be considered by automated spammer.

[1], had solution of only dictionary attack but not the other attack like online guessing attack, relay attack etc. are discussed in [7] also have limited success as one could not apply face as password everywhere.

In [12], access time increases due to applying Necklace graph in CAPTCHA. There is a certain lack of these views, on the current development trend of view in [13], about the application based on the shape feature,

5. CONCLUSION

In early days, an increasing number of public web services have made an attempt to prevent from exploitation by bots and automated scripts, by need of a user to solve a Turing-test issue commonly known as a CAPTCHA (Completely Automatic Public Turing test to tell Computers and Human Apart) before using the service. These efforts may be critical to the success of these sites in different ways. For example, Gmail enhance its service by blocking access to automated spammers, eBay improves its marketplace by blocking bots from flooding the site with scams, and Facebook limits creation of fraudulent profiles used to spam honest users or cheat at games. CAPTCHA has mainly two main characteristics: (a) be easy to solve by human and (b) be very hard to solve by computer script.

6. REFERENCES

[1] B. Pinkas and T. Sander, “Securing passwords against dictionary attacks,” in *Proc. ACM CCS*, 18-22 Nov. 2002, Washington DC USA, pp. 161–170.

[2] Moy G., Jones N., Harkless C., Potter R.,“Distortion estimation techniques in solving visual CAPTCHAs”, IEEE CVPR, pp. II-23-II-28, Vol. 2l, 2004

[3] Ahmad Salah El Ahmad, Jeff Yan,“Breaking Visual CAPTCHAs with Naïve Patter Recognition Algorithms”, IEEE Computer Security Applications Conference, Dec 2007, Miami Beach FL USA ,pp. 279- 291,.

[4] Jeff Yan, Ahmad Salah El Ahmad,"A low-cost attack on a Microsoft captcha", Proceeding of the 15th ACM Conference on Computer and communications security, October, 2008,Tacoma WA USA, pp. 543-554.

[5] Rich Gossweiler, Maryam Kamvar and Shumeet Baluja, "What's Up CAPTCHA? A CAPTCHA Based on Image Orientation", ACM 2009,20-24 April 2009, New York NY USA, pp. 841-850.

[6] R. Biddle, S. Chiasson, and P. C. van Oorschot, "Graphical passwords: Learning from the first twelve years," *ACM Comput. Surveys*, vol. 44, no. 4,4 Aug. 2012, New York NY USA.

[7] 2012, Feb. *The Science Behind Passfaces* [Online]. Available: http://www.realuser.com/published/ScienceBehindPassfaces.pdf

[8] Tsheten Tamang and Pattarasinee Bhattarakosol, "Uncover Impact Factors of Text-based CAPTCHA Identification", Computing and Convergence Technology (ICCCT), 2012 7th International Conference on IEEE, 13 June 2013,Seoul Korea(South).

[9] Chen-Chiung Hsieh and Zong-Yu Wu "Anti-SIFT Images Based CAPTCHA Using Versatile," IEEE, 2013.

[10] 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 "in IEEE transactions on information forensics and security, vol. 9, no. 6, june 2014, pp. 891-903.

[11] Anuj Thakur, Nikhil S., Rohit Chaware, SK Hafizul Islam, "The Reading Oriented Overlapping Text based CAPTCHA"ITACT-15 international conference in IEEE,21-22 Dec. 2015, Banglore India,pp.1-6.

[12] Mohamed Torky, Ali Meligy, Hani Ibrahim, "Securing Online Social Networks against Bad bots based on a Necklace CAPTCHA Approach.", 2016, IEEE 12th ICENCO,Giza Egypt, pp. 158-163.

[13] Gaihuan An, Wanjun Yu, "CAPTCHA Recognition Algorithm Based on the Relative Shape Context and Point Pattern Matching" 2017, 9th International Conference on Measuring Technology and Mechatronics Automation,14-15 Jan 2017,Changshu Hunan China, pp. 168-172.