A Review on Splitting and Rotating Captcha Technique

Mr.Parikshith Nayaka S K
Assistant professor , Dept. of CSE
Alva's Institute of Engineering & Technology
Mijar, Moodbidri, India

ShriRaksha
Student
Alva's Institute of Engineering & Technology
Mijar, Moodbidri, India

Sheetal
Student
Alva's Institute of Engineering & Technology
Mijar, Moodbidri, India

Vinutha
Student
Alva's Institute of Engineering & Technology
Mijar, Moodbidri, India

Sumana Rehman
Student
Alva's Institute of Engineering & Technology
Mijar, Moodbidri, India

Abstract— **Phishing is an attempt by an individual or a group to thieve personal confidential information such as passwords, credit card information etc from unsuspecting victims for identity theft, financial gain and other fraudulent activities. A simple username and password based authentication is not sufficient for web sites providing critical financial transactions. Where as a technique is needed for the websites to have a system which has the capability of differing human users and computer programs in reading images of text. Completely Automated Public Turing Test to Tell Computers and Human Apart (CAPTCHA) is such a defense system against Optical Character Recognition (OCR) software. OCR can be defined as a software which work for defeating CAPTCHA images and make countless number of registration on the websites. This study focuses on a new method which is splitting CAPTCHA images into several parts with random rotation values, and drawing random lines on a grid background. Lines are in the same color with the CAPTCHA text and they provide a distortion of image with grid background. In this paper, the algorithm of splitting and rotating method is introduced in detail.**

Keywords- **Phishing, Optical Character Recognition, image, CAPTCHA , Splitting & Rotating.**

I. INTRODUCTION

Most of the daily activities such as education, shopping or commerce are being carried out through the Internet. Users are commonly asked to fill out registration forms by entering required information to be able to operate specific tasks on the web sites. However, registration can be done by automated hacking software. Some people commit vandalistic acts such as attacking web sites with computer programs, and even can stop the running of the web site. These programs automatically fill out a form with wrong information to get in the web site. Therefore, web site holders are supposed to take precautions against those attacks for security.

Several defense systems have been proposed and presented in order to prevent such attacks. It is crucial for the websites to have a system which has the capability of distinguishing human users and computer programs in reading images of text. CAPTCHAs are challenge puzzles used to determine whether a user is human or not [1]. Intuitively, a CAPTCHA is a program

that can generate and grade tests that most humans can pass but current computer programs can not pass [2]. It stands for Completely Automated Public Turing Test to Tell Computers and Human Apart, and Public means that the code and the data used should be publicly available [3]. There are several types of testing such as pictures of objects, distorted text, or even audio clips for impaired users.

A more technical definition of CAPTCHA is provided in [4]: "CAPTCHA is a cryptographic protocol whose underlying hardness assumption is based on an AI problem". The most common applications for practical security by CAPTCHA test include online polls, free email services, shopping agents, search engine bots, worms and spam, and preventing dictionary attack [4]. For instance, email provider service es such as Hotmail and Yahoo provide a CAPTCHA test as a final step of the registration process to stop bots from subscribing and using their resources for spam distribution.

Turing test is used for providing the intelligence of a computer in the domain of Artificial Intelligence (AI). Turing tests use a method which put a human user and a computer in different rooms. There is also third room for the human interrogator to ask them questions. If the interrogator can not recognize the locations of human and computer, it results that the computer has passed the Turing test. CAPTCHA is a Turing test but it is quite different than the definition above. If the interrogator is replaced with computer rather than a human, then it is called as CAPTCHA. The main function of this method is human user can easily answer the interrogator's question but present computer programs are hardly or never can answer [5].

There are different types of CAPTCHA's. Image-based CAPTCHAs [6] require users to identify labeled images or rotated images. They evince a larger gap between human users and Bots. In implicit CAPTCHAs [7], the user does not have to read or type anything and just makes simple clicks on hot spots. Drawing captcha [8] generates numerous dots on a screen with noisy background, some of which are diverse from the others. Apart from the visual-CAPTCHAs, there exist a number of audio CAPTCHAs [9, 10, and 11] where the user must recognize and type the word that is played as a sound. In video-based

CAPTCHAs [12], users will be prompted to view a challenge video and then appropriately annotate (or tag) it.

One of the methods used in CAPTCHA is implementing the images of words. This method is based on the weak points of Optical Character Recognition (OCR) programs. OCRs are software to work for defeating CAPTCHA images and make countless number of registrations on the websites. OCRs can recognize the high quality texts using the common formats and standards [5].

This paper introduces a new CAPTCHA method which splits the image into several parts in the random width and height values. Additionally it rotates the split character in random rotation angles that yield a particular distortion in the image.   It is very difficult for OCRs to find out where characters are split and the end points of each image because of the random rotation. It would be very expensive to write an OCR algorithm to defeat our method CAPTCHA image is composed many images with random rotation values. The proposed method was implemented by the PHP (Personal Home Page) programming language. Section 2 introduces the previous studies on CAPTCHA. Design principles are presented in Section 3, and the details of our algorithm are explained in Section 4. Last section concludes the study, and shows the strengths and weaknesses of our method.

II.STUDIES ON CAPTCHA

A number of studies were conducted by researchers on developing new CAPTCHA. CAPTCHAs were originally developed by Alta Vista to avoid the submission of URLs to search engine Carnegie Mellon designed the Gimpy method which selects a word from dictionary and asks users to type what they see as an image after rendering the distorted image containing the text.

Yahoo uses the simple version of this method; EZ-Gimpy. EZ-Gimpy's image modification includes background grids, gradients, non-linear deformations, blurring, and pixel noise.

Figure 1. Some CAPTCHA words of Yahoo [9], Hot-mail[10] and G mail[11] respectively.

Hot-mail uses another CAPTCHA method in which a string of character is randomly selected. These are subjected to changes, later the users are asked to type what they see. There was a disadvantage in this method because some of the characters were read differently by putting curves between characters.Pix uses usual pictures instead of pictures of words.

However, this requires large space to store pictures.

There are two types of attack on CAPTCHA artificial intelligence and third party. To protect these attacks several enhancements are made. However, these enhancements are not without limitations. One such limitation is that most CAPTCHAs are not made with users that have impaired vision. There can also be technical limitations like CAPTCHA solution that require large amount of server storage, like those that depend on images, can increase implementation cost.

Poorly implemented CAPTCHAs can be broken easily even without using character recognition software. Some of the first generation CAPTCHAs has already been broken, so the new generation should be more powerful and complex to avoid from such attacks.

III.DESIGNING PRINCIPLES OF CAPTCHA

A fundamental requirement of CAPTCHAs necessitates that they must be designed to be easy for humans but difficult for computers i.e. the difficulty to be solved by OCR and any attack programs, readable common distortions, resisting malicious attacks, carrying many bits of information, the capability of coexisting with other CAPTCHAs, and little cognitive computation requirement by the user. The relative importance of these characteristics depends on the CAPTCHA type. The principles behind CAPTCHA are as follows:

1.The user is presented with a garbled image on which some text is displayed. This image is generated by the server using random text.

2.The user must enter the same letters in the text into a text field that is displayed on the form to protect.

3.When the form is submitted, the server checks if the text entered by the user matches the initial generated text. If it does, the transaction continues. Otherwise, an error message is displayed and the user has to enter a new code.

4.Exploits observation that humans are still much better than computers at many pattern recognition tasks.

IV.SUGGESTED ALGORITHM

In this paper, a new method has been developed for differing human users and computer programs from each other by mainly splitting CAPTCHA image into several parts with rotation and drawing a great deal of lines and circles randomly to the background. Additionally, a grid effect has been added to the background. Lines and circles have been randomly drawn in the color of text so that OCR program confuse while distinguishing which one is character or not.

CAPTCHA text consists of the characters and number in a range "ABDEFHKLMNPRSTUVWXZabdefgikmnopqrstuvwxyz0234 56789". The text is composed of five characters, and each character has its own bending and size value. Characters are split into several parts and each part is given randomly a rotation value in a certain angle domain interval such as: $[-1,1]-]$, $[3,3-]$, $[5,5$. Image parts are also split individually with random width and height values which provide an extra difficulty for OCR programs while finding the start and end of the images. Rotation in character parts provides confusion in recognizing the exact

one. The text shown in Figure 2 below is indeed'W9XZq'. This text is easily recognizable by the human but not OCR program. This CAPTCHA image is split into 8 parts as (4 X 2) matrix shape and each split has a random rotation angle value between -3 and3 degrees. Splits have random width and height values. Background and CAPTCHA text are in similar colors.
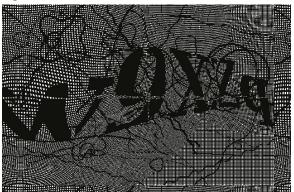

Figure 2.Drawing CAPTCHA image.

There is a grid in black color at the background. Lines and circles are also drawn in black color such as the CAPTCHA text. When you look at the first character, it is not easy to recognize the letter exactly due to rotation and splitting of character image. It seems like 'V' or 'I' or 'W'. In fact, it is 'W' but it is not recognizable for OCR program because character'W' is split into two different parts as 'V' and 'I'. The other letters in CAPTCHA image have same difficulty for OCR program. The programming steps of the algorithm that developed to generate CAPTCHA images are given with pseudo code and run-time output screen-shots as in follows:

Step 1. Start the session.

Step 2.Generate n letters random string from the string "ABDEFHKLMNPRSTUVWXZabdefgikmnopqrstuvwxyz023456789".// Take out some easy letters not to be confused by the user; C/G I/l Q/O h/b.// Users of this algorithm may choose other languages such as Arabic and Korean by modifying the string as they wish.

Step 3. Create the hash for the random text and put it into the session.

Step 4.Create transparent CAPTCHA image with w by h image size and add CAPTCHA text over it. Transparent CAPTCHA image with text can be created by specific PHP (Personal Home Page) built-in function: imagettftext()(imagettftext (resource $image , float $size , float $angle , int $x , int $y , int $color , string $fontfile , string $text)[6]).

Step 5. Set the initial X-position and Y-position of CAPTCHA image to 0.

Step 6. Split the CAPTCHA into k by l Matrix shape by dividing the CAPTCHA width into k parts and the height into l.

Step 7. Start a loop from 0 to k*l // After completing the first row in order to split into k parts, then pass to next row. If (i+1) Mod k+1 = 0 Then Set initial X-Position to 0 and initial Y-Position to Split Height (Image Height / l)End If.

Step 7.1. Create an array to put the split parts and put the split images into array.

Step 7.2. Randomize integer between -d and d to give random rotation to the splits. Rotate the splits with randomized variable that is random between -d and d.

Step 7.3. To pass to another split in one row, increase the initial X-Position by Split Width (Image Width / k)in each loop step.

Step 7.4. End Loop.

Step 8. Combine the splits to create new CAPTCHA with split and rotation.

Step 9. Add background to transparent new CAPTCHA image object with randomly drawn lines and special effects (Number of lines=250, line color is black and add grid effect).

Step 10. Export the final CAPTCHA image as a JPEG file in the name of 'captcha.jpg'.

Step 11. Destroy the final CAPTCHA image object to be refreshed in each session.

V.CONCLUSION

The concept behind in this field started from the problems faced during online transactions. So CAPTCHA was additionally added to test whether user is human. By using this method it is safer since OCR programs cannot read it. This method provides better security in terms of not allowing the intruder log in into the account even when the user knows the username of a particular user. Finally it needs less processing requirements and can be operated in small size of bandwidth. This technique can be improvised using other powerful algorithms like super CAPTCHA and also using RECAPTCHA technique.

VI.REFERENCES

[1] Blum, M., 2000, The CAPCTHA Project, Completely Automatic Public Turing Test to Tell Computers and Humans Apart", Dept. of Computer Science, Carnegie-Mellon University, http://www.captcha.net

[2] Athanasopoulos, E., Antonatos, S., "Enchanced CAPTCHAs: Using Animation to Tell Humans and Computers Apart", LNCS, 4237,2006, pp. 97-108.

[3] Wang, S., Baird, H., Bentley, J., "CAPTCHA Challenge Tradeoffs: Familarity of Strings versus Degradation of Images", The 18th International Conference on Pattern Recognition, ICPR'06, IEEE, 2006.

[4] Von Ahn, L., Blum, M., Nicholas, J.H., Langford, J., "CAPTCHA: Using Hard AI Problems For Security", In Proceedings of Eurocrypt, pp.294-311, 2003.

[5] Shahreza, M., Shahreza, S., "Preventing Mobile Software Cracking Software", IEEE, Innovations in Information Technology, Dubai, 2006, pp. 1-5.

[6]Merler, M. & Jacob, J., (2009) "Breaking an Image based CAPTCHA", Technical Paper submitted to the Department of

Special Issue - 2019

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
RTESIT - 2019 Conference Proceedings

Computer Science, Columbia University, USA, Spring term.

[7] Baird, H.S. & Bentley, J.L., (2005) "Implicit CAPTCHAs", Proceedings of the SPIE/IS&T Conference on Document Recognition and Retrieval XII (DR&R2005), San Jose, pp. 191-196.

[8] Shirali-Shahreza, M. & Shirali-Shahreza, S., (2006) "Drawing CAPTCHA", Proceedings of the 28th International Conference Information Technology Interfaces (ITI 2006), Cavtat, Dubrovnik, Croatia, June 19-22, 2006, pp 475-480.

[9] Kochanski, G. et al., (2002) "A Reverse Turing Test using speech", Proceedings of the Seventh International Conference on Spoken Language Processing (ICSLP2002 - INTERSPEECH 2002), Denver, Colorado, USA, September 16-20, 2002, pp 1357-1360.

[10] Chan., T.Y. (2003) "Using a Text-to-Speech Synthesizer to Generate a Reverse Turing Test," Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence, 2003, pp 226-232.

[11] Schlaikjer, A., (2007) "A Dual-Use Speech CAPTCHA: Aiding Visually Impaired Web Users while Providing Transcriptions of Audio Streams", Technical Report CMU-LTI-07-014, Carnegie Mellon University, November.

[12] Kluever, K.A., (2008) "Evaluating the Usability and Security of a Video CAPTCHA", Master's thesis submitted to Rochester Institute of Technology, Rochester, New York, August.