Puzzle Based Captcha Implementation for Noisy Environment

Kanwaldeep Kaur Kanwal[1], Anupama Gupta[2], Vivek Aggarwal[3*], Amandeep Kaur[4]
LLRIET Moga-142001 (Punjab), India
kanwal.kammo@gmail.com, anupamagemini@gmail.com, *agarwalz_v@yahoo.com,
amannagpal.kaur@gmail.com

***Abstract*—** *Today, it is a very common problem that bots attack on the online polls and register free email accounts automatically that increase the congestion on network as well as consume large amount of server space. Therefore, to prevent these kinds of attacks, a technique called Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) have been used. The main motive of present research work is to design a CAPTCHA in order to increase the security by preventing bot attacks using random mathematical functions and background noise makes it invulnerable for the optical recognition (OCR) technique to break the CAPTCHA as OCR attack is capable of only extracting characters from an image. To make CAPTCHA more secure, cross operations has been embedded in the present algorithm which makes it impossible for OCR technique to decode its output result.*

Keywords*— *CAPTCHA, OCR, Puzzle, Cross operation, Bots.

I. INTRODUCTION

CAPTCHA is an acronym for Completely Automated Public Turing test to tell Computers and Humans Apart. The term "CAPTCHA" was coined in 2000 by Luis Von Ahn [1], Manuel Blum, Nicholas J. Hopper (all of Carnegie Mellon University, and John Langford (then of IBM). They are challenge-response tests to ensure that the users are human not a bot. The use of a CAPTCHA is to block submissions of spam bots that can be harvest email addresses from publicly. A very common kind of CAPTCHA which is used on websites requires the users to enter the string of distorted string of characters on the screen. CAPTCHAs are used because it is difficult for the computers to extract or understand the distorted text or image, whereas it is relatively easy for a human to understand that text or image. Therefore, the correct response to a CAPTCHA challenge is assumed to come from a human and the user is permitted into the website.

The need for CAPTCHAs rose to keep out the website or search engine from bots. In 1997, AltaVista developed a method to generate a printed text randomly that only humans could read and not machine readersThe existing CAPTCHAs can be generally classified into three categories: Image-based CAPTCHAs [2, 3], Text-based CAPTCHAs [4, 5] and Sound-based CAPTCHAs [6]. Text-based CAPTCHAs that depend on the distortion of digits,letters and other visual effects added in the background image. The user is asked to identify the distorted characters and entered them. So far, most commonly used CAPTCHAs are text-based CAPTCHAs. They can be easily designed and implemented and can be easily solved by users.. Examples of text and Graphic CAPTCHAs include:

EZGimpy: Pick a word or words from a small dictionary. Distort them, add noise and background.

Gimpy-r: Pick random letters. Distort them, add noise and background.

Baffle Text: Pick random Alphabets which create nonsense but pronounceable words.

Bongo: User has to solve pattern reorganization problem by finding which figure belongs to which one.

Pix: User has to recognize the common features from a set of images.

The organization of the paper trails as: Review of previous related work is given in Section. II. Section III focuses on the formulation of the proposed algorithm. Section IV reports a number of experimental results to demonstrate the performance of the new algorithm. Finally, conclusions are drawn in Section. V.

II. LITERATURE SURVEY

Castro et al. (2009) presented a black-box attack. Its aims to protect a free service delivered using the Internet. This CAPTCHA was referred to as ‘‘QRBGS CAPTCHA’’ or ‘‘Math CAPTCHA’’. In this it gave the user mathematical problems to solve in order to prove human. This required no development in Artificial Intelligence or automatic character recognition, the intended path, thus becoming a side-channel attack, based on the previously mentioned CAPTCHAs flaws. The author concluded with some tips for enhancing this CAPTCHA that can be considered as general guidelines [6]. Biddle

et al. (2011) stated as many graphical password schemes have been proposed as alternatives to text-based password authentication and yet there still remains research to be carried out in this field. So, the author provides a comprehensive overview of published research in the area, covering usability and security aspects, as well as system evaluation. The paper first catalogues existing approaches, highlighting novel features of selected schemes and identifying key usability or security advantages. The author then review usability requirements for knowledge-based authentication as they apply to graphical passwords, identify security threats that such systems must address and review known attacks, discuss methodological issues related to empirical evaluation, and identify areas for further research and improved methodology [7]. Choudhary and Saroha (2013) stated that the purpose of a CAPTCHA is to block form submissions from spam bots – automated scripts that harvest email addresses from publicly available web forms. The term "CAPTCHA" was coined in 2000 by Luis Von Ahn, Manuel Blum, Nicholas J. Hopper (all of Carnegie Mellon University, and John Langford (then of IBM). CAPTCHAs are used because of the fact that it is difficult for the computers to extract the text from such a distorted image, whereas it is relatively easy for a human to understand the text hidden behind the distortions. Therefore, the correct response to a CAPTCHA challenge is assumed to come from a human and the user is permitted into the website. The CAPTCHA test helps identify which users are real human beings and which ones are computer programs[8]. Parveen and Singh (2014) stated CAPTCHA‟s were used to improve the security of Internet based applications in order to ensure that a web based application which was intended to be used by a human being is not maliciously used by Artificially Intelligent programs called bots. As the current CAPTCHA methods are striving to turn out to be difficult for bots, they are gradually becoming difficult and annoying for human users as well. This paper carries out a systematic study of various Text-based CAPTCHA‟s and proposes the application of Forepart based prediction and Row-wise mapping to break these CAPTCHA‟s to evaluate their robustness. CAPTCHA segmentation and recognition is based on Forepart prediction, necessity sufficiency matching and masking [9]. Yan et al. (2008) suggested CAPTCHA is now almost a standard security technology, and has found widespread application in commercial websites. Usability and robustness were two fundamental issues with CAPTCHA, and they often interconnect with each other. This paper contains usability issues that should be considered and addressed in the design of CAPTCHAs. Some of these issues were intuitive, but some others have subtle implications for robustness (or security). A simple but novel framework for examining CAPTCHA usability is also proposed by the authors [10]. Ali and Karim (2014) described CAPTCHA as a technique of testing for ensures that only people just passing the test and not the system computer generated (bots). The development of CAPTCHA system is to provide creative and validation tests that can be easily solved by humans and difficult for bots. There are four types of methods in development CAPTCHA which is the text-based CAPTCHA; CAPTCHA based on image, CAPTCHA based audio and video-based CAPTCHA. CAPTCHA system based on puzzle is developed by using a technique based on an image CAPTCHA. Then, this type of CAPTCHA was developed by using HTML, JavaScript/J-Query and Cascading Style Sheets (CSS). The CAPTCHA system was developed using a sequence of phases of development that are in the methodology of evolutionary prototyping model. Applications certification is expected a confirmatory test that is easily solved by the user and then be interactive authentication applications [11].

III. PROPOSED WORK

In this research, a technique has been designed to fortify web security and to prevent bots from damaging or entering the system.

Step 1: Generate a noisy background for the captcha image so that it cannot be read easily by the BOT or OCR techniques easily.

Step 2: Generate different random mathematical expressions. These mathematical expressions include series equation, measurement related functions, work, time, speed, distance and cross operations as puzzles.

Step 3: Generate random letters of different font size, font style and orientation with noise.

Step 4: Create the captcha database file.

Step 5: Check the performance by applying OCR technique.

In the cross operation the meaning of the operators are switched. For example: if human performing an operation in the mathematics we denote add by multiply, multiply by add, divide by subtraction and subtraction by divide in the given expression. Thus the original operation is not visible to the front end and the BOT recognize it as the default operation as per the captcha image and not breakable by the BOT or OCR techniques. Whereas, for the user understanding we provide hints

about the operation so that human being can solve it easily. For instance, the actual meaning of Fig. 1 is shown in the below mentioned equation:

28*4+7*4/10 = 14

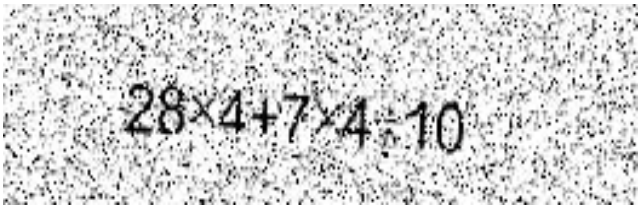

FIG. 1 CROSS OPERATION

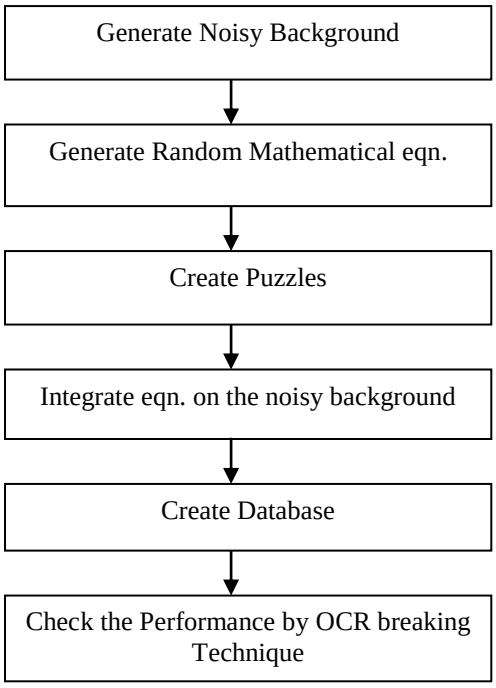

FIG. 2 FLOW CHART OF PROPOSED WORK.

IV. RESULTS AND DISCUSSION

This experiment is carried out using MATAB. The captcha designed in this research contains mathematical formulae which could not be cracked by the OCR. The proposed CAPTCHA has been tested by many users and we have concluded that only humans were capable of solving our proposed mathematical image CAPTCHA and the results are shown below in Table 4.1. Thus the proposed CAPTCHA in this research is effective and secured and more simple as compare to first one. The Table 1 shows the success rate and response time. The success rate is calculated in terms of successful attempt made by the user in recognizing and attaining correct answer for the equations shown in the CAPTCHA image. It is calculated by number of correct qualified CAPTCHA divided by total number of passed CAPTCHA multiplied by 100. On the other side, the response time represents the time taken by the user in solving the equation and it calculated by using TIC-TOC command of MATLAB.

The Register window contains a mathematical equation in the form of a CAPTCHA image embedded with puzzles and to register the user must solve the equation by solving the values of puzzle window in which meaning of the operators are also shown via Fig 3.

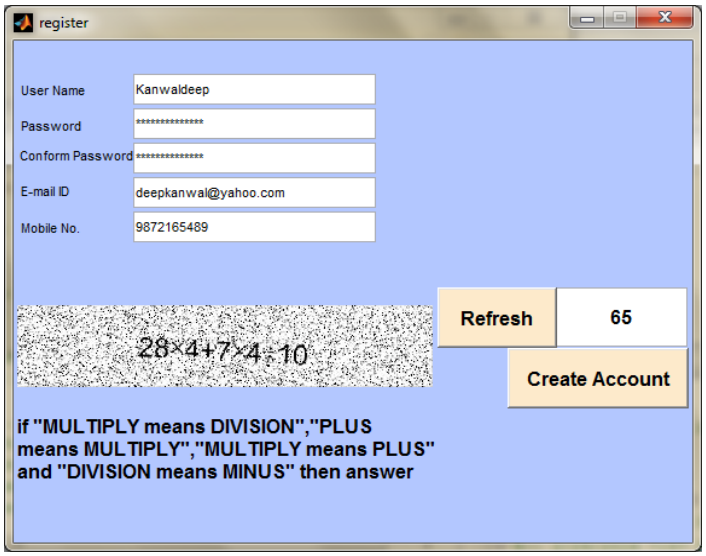

FIG. 3 USER REGISTER WINDOW WITH CAPTCHA

Table 1 show the values of success rate and response time of captcha

TABLE 1
RESPONSE TIME AND SUCCESS RATE

S.No	Success Rate (in %)	Response Time (in s)
1.	85	5
2.	90	8
3.	88	9
4.	89	6

The Break CAPTCHA window enables the user to Load any CAPTCHA image containing mathematical equation and by pushing "OCR Attack" button the algorithm then performs the attack and displays the extracted character from the loaded CAPTCHA image. As seen from the Fig. 4 the extracted character does not match with the loaded CAPTCHA image.

As shown in Fig. 5 the OCR attack was able to extract the characters from the noisy captcha image. This proves the previous CAPTCHA algorithm to be unworthy.

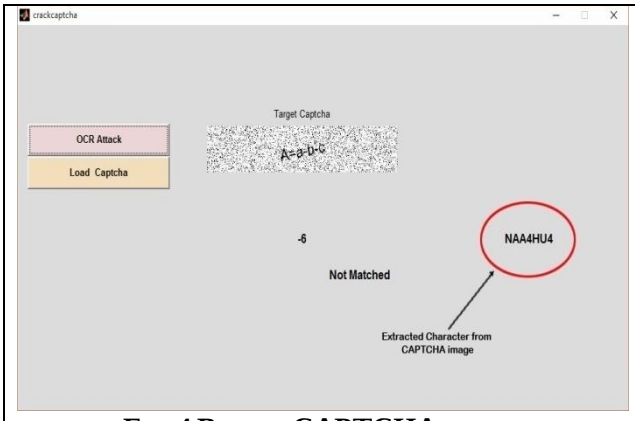

FIG.4 BREAK CAPTCHA WINDOW

FIG. 5 CRACKED CAPTCHA ALGORITHM

FIG. 6 PROPOSED INVULNERABLE CAPTCHA

And the loaded CAPTCH image in Fig. 5 and 6 contains the following equation:

$$A = a^3 \quad (1)$$

And the characters extracted by the OCR attacks are "AQ43". This proves the proposed algorithm to be resilient against OCR attacks. Thus it is clear from the OCR Attack results that the CAPTCHA proposed in this research is effective and cannot be cracked by bot because it contains some mathematical equations as puzzle. However, if in case the OCR attack is able to extract the exact information but still it won't be able to crack the equation as OCR considers its own output as string of characters and not as an equation.

V. CONCLUSIONS AND FUTURE SCOPE

In this research a new captcha design approach has been implemented using simple mathematical equations and puzzles. The result shows that the proposed approach is more secure and simple as compare to others. Furthermore, after applying OCR attacks user can conclude that our proposed method is more secure to bot attacks and is very useful for high level of security programs.

This research work can be further extended or enhance by including other techniques along with math captcha with Boolean expressions, number system to make it more rebust and unbreakable.

REFERENCES

[1] Ahn, L. V., Blum, M., and Langford, J. (2003). CAPTCHA: Using hard AI problems for security. IBM T.J. Watson Research Center .

[2] Ahn, L. V., Blum, M., and Langford, J. (2004). Telling human and computers apart automatically. Communications of the ACM , pp. 57-60.

[3] Elson, J., Douceur, J. R., Howell, J., and Saul, J. (2007). Assira: a CAPTCHA that exploits interest-aligned manual image categorization. Proceedings of the 14th ACM conference on Computer and Communications Security , pp. 366-374.

[4] Chew, M., and Baird, H. S. (2003). Baffle Text: A Human Interactive Proof. Proceedings of the 10th SPIE/IS&T Document Recognition and Retrieval Conference .

[5] Coates, A. L., Fateman, R. J., and Baird, H. S. (2001). Pessimal Print: A Reverse Turing Test. Proceedings of the 6th International Conference on Document Analysis and Recognition , pp. 1154-1158.

[6] Castro, C. J. (2009). Pitsfall in CAPTCHA design and implementation: The Math CAPTCHA, a casr study. ELSEVIER , pp. 141-157.

[7] Biddle, R., Chiasson, S., & Oorschot van, P. C. (2011). Graphical Passords: Learning from the First Twelve Years. School of Computer Science, Carleton University.

[8] Choudhary, S., Saroha, R., Dahiya, Y., & Choudhary, S. (2013). Understanding CAPTCHA text and audio based CAPTCHA with its applications. International Journal of Advanced Research in Computer Science and Software Engineering , 3 (6).

[9] Parveen, H., & Singh, S. (2014). CAPTCHA recognition and robustness measurement using hybrid approaches. International Journal of Advanced Research in Computer Science and Software Engineering, 4 (6).

[10] Yan, J., & Ahmad, S. E. (2008). Usability of CAPTCHAs or usability issues in CAPTCHA design. SOUPS

[11] Ali, B. H., & Karim, F. B. (2014). Development of CAPTCHA system based on puzzle. International Conference on Computer, Communications and Control Technology , pp. 426-428.