SD-CAPTCHA: An Alternate to Authentication System

Mrs.. Sumeet Kaur, *Asst. Prof. Department of Computer Engineering, YCOE*
Mr. Puneet Kumar Bansal, *Student, M.Tech C.E, YCOE*

Abstract*: Captcha stands for completely automated program to tell computer and human are apart. This captcha system is not a new concept it's being used by number of online service providers such as Google, Yahoo etc. but these captcha are highly sophisticated and sometime user gets annoyed to get rid of it. The prime task of Captcha is to differentiate computer from human and with the advent of technology and attacks the captchas are getting more secure and stronger day by day, with time usability or friendliness is reduced leaving user annoyed because machine passes the captcha test but user yet messing up with the captcha to pass. In this paper a standalone developed captcha design is proposed which is not only secured but also it is user friendly. A basic attack was made to check the resistance of SD-CAPTCHA which gives results lesser then 55%, because the robustness of a text CAPTCHA is typically determined by the strength of its segmentation-resistance mechanism.*

Keywords: *CAPTCHA, Authentication, usability, security, robustness, anti –segmentation, anti-recognition.*

1. INTRODUCTION

CAPTCHA (Completely **A**utomated **P**ublic **T**uring test to tell **C**omputer and **H**uman **A**part) was deployed to improve the security of server service and verify that a client request is submitted by individual users from online operations rather than by malicious software. It is a program that generates and grades tests that humans can pass easily, whereas computers cannot [1]. A good CAPTCHA should satisfy two main aspects: robustness and usability. The robustness is its strength to defend against adversarial attacks; whereas the usability is the ease with which humans pass its challenges [5]. Over the past decade, the generation of CAPTCHAs has been developed through a number of varied distortion mechanisms, for example, random arcs and random angled connecting lines. Although CAPTCHA designers attempt to satisfy both robustness and usability, currently deployed CAPTCHAs mostly concentrate on robustness due to an increase in the number of attacks on the previous designs [11,12]. Whereas, usability has been studied, mainly, on a functional level, with focus on differences in expected accuracy and response time [2,3,4,7], and on systematic analysis of usability issues that should be considered in the design [5].

There are three main types of CAPTCHA, text-based, sound-based and image-based. Text-based CAPTCHA with many advantages is the most commonly deployed type in websites to date, [2]. Text-based schemes prompt users to recognize text, which state-of-the-art text recognition programs cannot perform. In this paper, the term CAPTCHA refers to text-based schemes only because of popularity and ease of use the focus of this paper is in text based schemes only..

2. RELATED WORK

A considerable number of studies have been conducted by researchers in CAPTCHA usability. C. Kumar, L. Kevin, S. Patrice Y. and C. Mary in [2] discussed several CAPTCHA generators with a user-friendly design along with an examination of the effect of different text distortion techniques on the readability of a CAPTCHA designed by Microsoft. B. Henry S, M. Michael A. and W. Sui-Yu, and C. Monica and B. Henry S in [8,9] respectively discussed the accuracy of the readability issues of CAPTCHA.

B. Henry S and R. Terry in [6] discussed the legibility of Scatter Type CAPTCHA demonstrating some confusion between pairs of characters, and suggesting the construction of classifiers for legibility. Y. Jeff and A. El Ahmad in [5] discussed the usability issues that should be considered and addressed in the design of CAPTCHAs, such as content issues. Consequently, a simple but novel framework was proposed to examine CAPTCHA usability. However, it did not discuss how to

improve the usability issues related to character confusion. B. Elie, B. Steven, F. Celine, M. John C. and J. Dan performed a large scale study in [7] assessed how well CAPTCHAs achieve the requirement of making it easy for humans to pass the test. It demonstrated that CAPTCHAs are quite difficult for humans to read. L. Ying-Lien and H. Chih-Hsiang in thier recent study in [14] conducted an experiment to study the effect of age groups and distortion types on the CAPTCHA task. This study has significant implications for the design of CAPTCHA due to the inevitability of the alternative security measure and the increasing population of elder internet users. The robustness of a CAPTCHA was attracted substantial attention in the research community, e.g. in [11,12,13]. These discussions conclude that a CAPTCHA should be segmentation-resistance rather than recognition-resistance.

3. CAPTCHA DEVELOPMENT SYSTEM

CAPTCHAs are considered as a standard defense against malicious Internet programs. However, the number of attacks on CAPTCHAs has seen a continuous increase. This leads to the development of new designs. Developing a new design should not only learn from previous attacks, but should also apply security engineering knowledge to the design. Thus, CAPTCHA developing system is introduced to show the targeted area that satisfies the security and usability aspects, as shown in Fig. 1.

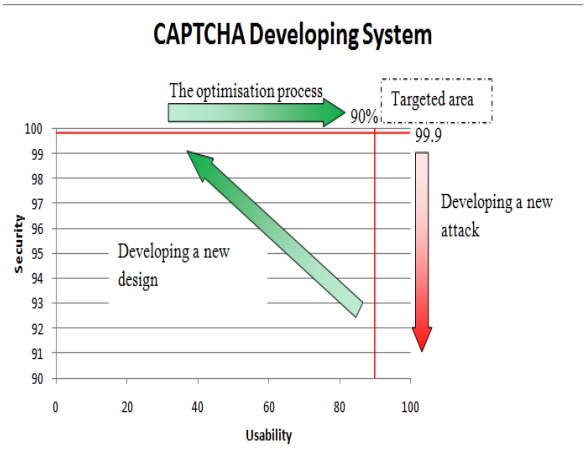

Fig 1. The Captcha Development System [15]

The above section highlight the kinds of attacks that CAPTCHAs have to defeat, developing a new design with regards to security and usability aspects and the importance of the optimization process.

A. DEVELOPING A NEW ATTACK

The reasonable success rate of an attack on a CAPTCHA is given by [2] as a rule: computer programs should not have a success rate higher than 0.01%. In spite of this, the success rate of some developed attacks reaches higher rates, either by using sophisticated object recognition algorithms [13] or by using naive pattern recognition algorithms [11]. Furthermore, the state of the art emphasizes that the robustness of CAPTCHAs should not only rely on which character it is, but also on the difficulty of finding where the character is (Segmentation). Although deployed segmentation resistance mechanisms can present such robustness, it can be vulnerable to simple but novel attacks such as in [12].

B. DEVELOPING A NEW DESIGN

Developing a new attack that will crack the currently deployed CAPTCHA challenges is not an easy task but understanding the design of CAPTCHA with regards to security aspects and also the usability aspects, the human success rate should approach at least 90% [2].

4. PROPOSED SCHEME

As discussed in the previous sections, segmentation-resistance plays an important role in providing appropriate security guarantee. We propose a new scheme named SD-CAPTCHA i.e. Standalone Captcha in which captcha is designed as a standalone captcha which generates such challenges that are hard to segment. The robustness of a text CAPTCHA is typically determined by the strength of its segmentation-resistance mechanism. This proposed scheme generates captcha challenges using fixed set of characters and applied different anti-segmentation and anti-recognition techniques deployed by different authors at regular interval of time. Fig 2. shows some of the

generated captcha challenges from SD-CAPTCHA.

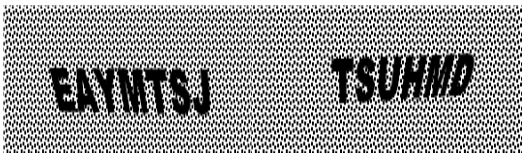

Fig 2. SD-CAPTCHA generated captcha challenges

A. **CHARACTERISTICS OF ROPOSED SCHEME**:

The following features has been considered while designing the proposed SD-CAPTCHA

1. Uppercase English Letters are used at random for enhanced usability
2. Length of the captcha challenge will be between 4 to 7 so as to have bit randomization.
3. Foreground color is black only and Background color is black & white so as to resist the binarisation attack.
4. Random Angle between -30^{o} to 30^{o} to the text of the captcha challenge.
5. Auto shrinks & auto expansion of the text of the challenge depending on the length & width of text to keep it inside the fix boundary dimensions.
6. All characters of the captcha challenge are kept bold to keep challenge user friendly and to resist pixel count attack.

B. PERFORMANCE ANALYSIS

The biggest challenge while designing a captcha is to keep it simple for user so that human can easily pass the test but yet difficult for the computer. The designed captcha is strong only if it resists a certain attack and the best part is the developer knows the loop holes of the design and the attack designed should be such that it breaks currently deployed except google, yahoo, MSN etc captcha but the designed captcha could not get break down using any crack. The only way to analyze the performance of SD-CAPTCHA is to pass through an attack

The performance of the SD-CAPTCHA is measure using the formula:

$$\text{P.S.A} = \frac{\text{Number of Segments Produced}}{\text{Length of SD-CAPTCHA}} \text{ X } 100$$

Note: P.S.A is Probable Solving Accuracy

C. **ATTACK MADE ON SD-CAPTCHA:** The attack that we have applied on the SD-CAPTCHA is comprised of the following steps:

1. Read the image from standard database.
2. Convert the image into black & white
3. Compliment the image.
4. Remove the pixels or mark them 0 whose value is less than 125.
5. Compliment the image.
6. Remove the pixels whose value is less than 50 so as we left with the dark pixels only.
7. Select the bounding regions of the image and assign them a fix color.
8. Crop the characters and stores it into the database.
9. Count the number of segments that has successfully cropped.

This particular attack can break any captcha challenge that has some space between each character of the challenge and not connected with each through any mean.

5. RESULTS

In order to demonstrate the success of the proposed SD-CAPTCHA, a CAPTCHA generator is developed. This generator can produce challenges with different types of challenge text. These types include random strings of different length. The naïve attack was applied on 50 challenges generated from SD-CAPTCHA and 50 samples collected from various websites using captcha system also comprising the EZ-Gimpy captcha which sought to be tough to be crack but actually not. The results obtained after applying designed naïve attack are as follows:

Sr No	Characteristics	Sample Set	Designed Set
1	Average Turn Around Time (µs)	7.29	6.83
2	Average Number of Segments	5.28	2.92
3	Probable Solving Accuracy (%)	105	53

Table 1.1 Result Analysis

Discussions: The above depicted results in the table 1.1 shows the following results

1. The Average Turn Around time depicts the time taken by the breaking process of the attack applied and it is measured in microseconds (µs).

2. The Average number of Segments created after applying attack on each image irrespective of the length of the captcha challenge and it is measured in numbers.
3. The probable solving accuracy depicts that the challenge generated has succeeded in segmenting only 53% of the characters in the challenge rest of the characters are un-segmentable. It is clear that if we do not get the equal number of segments then it is not possible to break it even with highly sophisticated OCR Programs and it is measured in percentage (%).

6. FUTURE WORK

The improvement over captcha design and keeping it user friendly and yet making it secure for using it as online authentication system offers a great opportunity to its wide adoption on the real time websites such payment gateways, online registration websites etc. We envision future improvements to be made in the following aspects:

- Randomization operations of alterations on each individual character of the string of the text based captcha.
- Randomization of the background image of the captcha.

7. CONCLUSION

The results obtained clearly depicts that the current text based captchas that are being used by different websites are not secure enough because with very simple attack made on sample captcha collected from lie websites gives 90% of the captcha challenges are breakable and giving 100% results. But the captcha that we have designed is more secure and more user friendly than the currently text based captcha deployed. It is concluded that every captcha designed by me is secure and yet user friendly giving only 53% accuracy in segmenting the captcha Challenge and hence the computer program attack that I tried giving results less than 0.01% because it is not possible or worth trying the recognition phase.

8. REFRENCES

[1] Von Ahn Luis, B. Manuel and L. John,"Telling Humans and Computer Apart Automatically", CACM, V47, Issue 2, Pages 56-60.

[2] C. Kumar, L. Kevin, S. Patrice Y. and C. Mary, "Designing Human Friendly Human Interaction Proofs," ACM CHI'05, Pages 711-720

[3] C. Kumar, L. Kevin, S. Patrice Y. and C. Mary, "Building Segmentation Based Human friendly Human Interaction Proofs," 2nd Int'l Workshop on Human Interaction Proofs, Springer- Verlag, LNCS 3517, pages 1-26.

[4] W. Sui-Yu, B. Henry S. and B. Jon L.,"CAPTCHA Challenge Tradeoffs: Familiarity of Strings Versus Degradation of Images," In ICPR '06, Volume 3, Pages 164-167.

[5] Y. Jeff and A. El Ahmad, "Usability of CAPTCHAs or Usability Issues in CAPTCHA Design," In SOUPS '08, Pages 44–52

[6] B. Henry S and R. Terry, "Scattertype: A Reading CAPTCHA Resistant to Segmentation Attack," In ICDAR' 05, Pages 935-939

[7] B. Elie, B. Steven, F. Celine, M. John C. and J. Dan, "How Good Are Humans at Solving CAPTCHAs? A Large Scale Evaluation," In IEEE S&P '10, Pages 399-413.

[8] B. Henry S, M. Michael A. and W. Sui-Yu, "A Highly Legible CAPTCHA That Resists Segmentation Attacks," in HIP'05, Springer-Verlag. LNCS 3517, Pages 27-41

[9] C. Monica and B. Henry S, "BaffleText: a human interactive proof," Proc. of 10th IS&T/SPIE Document Recognition & Retrieval Conference, Pages 22-23.

[10] W. Jonathan, *Strong CAPTCHA Guidelines: v1.2*. http://bitland.net/ captcha.pdf, December.

[11] Y. Jeff and A. El Ahmad, ''Breaking Visual CAPTCHAs with Naïve Pattern Recognition Algorithms,'' In *Proc. Of the 23rd Annual Computer Security Applications Conference (ACSAC'07),* FL, USA, Dec. 2007, IEEE computer society, Pages. 279-291.

[12] Y. Jeff and A. El Ahmad, ''A Low-cost Attack on a Microsoft CAPTCHA,'' In 15th ACM Conference on Computer and Communications Security (CCS'08). Virginia, USA, Oct. 2008, ACM Press. pages. 543-554.

[13] M. Greg and M. Jitendra, "Recognising Objects in Adversarial Clutter: Breaking a Visual CAPTCHA," In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03), vol. 1, June. 2003, Pages.134-141.

[14] L. Ying-Lien and H. Chih-Hsiang, "Usability Study Of Text-based CAPTCHAs," Displays, vol. 32, April. 2011, Pages. 81-86.

[15] Suliman A. Alsuhibany, "Optimising Captcha Generation", 2011, Sixth International Conference on Availability, Reliability and Security of IEEE Computer Society, Pages 740-745.