



Analysis of Different Text Based Captcha Methods

S.Thangavelu¹; T.Purusothaman²

¹M.S.University, Tirunelveli, India

²Govt. College of Technology, Coimbatore, India

¹Email: s_thangavelu@rediff.com

Abstract— *CAPTCHA stands for completely automated public turing test to tell computers and humans apart. The primary duty of Captcha is to separate humans and bots. Captcha generates a challenge response test and direct the users to perform a simple task. It is easy for the humans to complete the task, whereas it is difficult for the bot programs. Thus Captcha becomes a standard security mechanism in all the websites and web applications to prevent Internet Bots from unauthorised access.*

Keywords— *Captcha, bots, Security, text, image, audio, video*

I. INTRODUCTION

Captcha also called as Human interactive proof engaged in all the websites and web applications which distinguish humans and computer bots automatically. The word Captcha was devised by Luis von Ahn [1], Manuel Blum, Nicholas J. Hopper, and John Langford of Carnegie Mellon University during 2003. The various types [2] of Captcha methods are,

- Text based Captcha methods
- Image based Captcha methods
- Video based Captcha methods
- Audio based Captcha methods

II. CAPTCHA TYPES

Text-based Captcha methods: The text based Captcha [3] methods are the most common and universally accepted form of Captcha. In the text based Captcha methods a sequence of characters are randomly generated, distorted and presented to the users. Also to make the Captcha test difficult for the bots, noise added in the background. The user is asked to type the characters in the input text box for validation. The user is granted access only when the input characters matches with the Captcha characters. Text based Captcha are very user friendly when compare to others. Lot of improvements and alterations are made in the text based Captcha methods time to time and implemented in all the popular websites as a security arrangement.

Image based Captcha methods: Image based Captcha methods are also employed in most of the websites. In this method the user is required to perform some image recognition task [4]. Humans have the capability to recognize different types of images, patterns etc. In the image based Captcha method the user have to identify

the images and need to perform different kind of actions such as solving a quiz, matching the symbols, recognizing the human faces etc.

The image based Captcha are difficult for generation and also it is difficult for the users having low vision and color blindness. It also suffers by machine learning and edge detection attacks.

Audio based Captcha methods: Audio Captchas are based on sound systems [5]. The audio based Captcha was initially created for the benefit of visually disabled users. Audio based Captcha uses downloadable audio clippings. In this method the program picks a word or a sequence of numbers at random and turns into sound clip and distorts the sound clip. The users need to enter the correct word for validation. This Captcha test is based on the understanding capacity of human to the words and the accent which is a very tough task for bots. The audio based Captcha are difficult for users with hearing impairments. Also, the users need sound knowledge in English.

Video Based Captcha methods: In the Video based Captcha methods a video clipping which contains few key words are presented to the user. Some websites uses a video from the public domain like YouTube as Captcha. The users need to identify the words in the video and submit them for validation. This method provides better security and enhanced usability.

The video files are large in size when compare to text Captcha. Large file size needs extra download time or require high bandwidth. Also the users need more concentration to identify the words.

III. APPLICATIONS OF CAPTCHA

Free E Mail Services: A large number of websites on internet offer free registration to the Social networking sites, Email services, Web blogs etc. These web sites are attacked by the internet bots which signup automatically in huge amount continuously. This intern causes denial of service to the humans on the particular web service. Captcha prevent such attacks.

Web crawler: WebCrawler or Spider is a program used by search engines to explore the Internet and automatically download web content available on web sites. They capture the text of the pages and the links on the pages, thus it enables search engine users to find new pages. The bots acquire specific types of information from the web pages such as mining email addresses which can be used mostly for spam.

Dictionary Attacks: Captcha prevents dictionary attacks on passwords [7]. Bots through iteration method try to break the passwords of the users. Introducing Captcha in the signup page prevents such attacks. Humans need to type the passwords along with Captcha.

Online Games: Captcha prevents web Robots from playing online games

Phishing Attacks: Phishing attack is a novel attack in which the bots are attempting to steal the user credentials such as bank account details, usernames, passwords, credit and debit card details; Captcha prevents any such phishing attacks.

IV. ANALYSIS OF DIFFERENT TEXT BASED CAPTCHA METHODS

Text based Captcha methods are one of the most common and widely accepted forms of Captcha. The Captcha implementation was first developed by Andrei broder in 1997 for AltaVista website; to avoid automatic adding of URLs to the website for indexing. Thereafter many varieties Captcha methods are introduced for web security.

Gimpy Captcha: Gimpy is one of the trustworthy text based Captcha in the early days, developed by Carnegie Mellon University in association with Yahoo. Gimpy Captcha [6] is based on the human ability to read the extremely distorted text. The internet bots were unable to identify the Gimpy Captcha for a long time. Gimpy indiscriminately selects several words from a dictionary and displaying them with different direction, distortion and overlapping. The client is asked to enter a subset of the words in the image for authentication. Humans only can identify the words correctly, while the automated bot programs cannot. The Gimpy Captcha image is shown in Figure.1.



Figure.1. Gimpy Captcha

EzGimpy Captcha: EzGimpy is another version of Gimpy Captcha designed in simple form. EzGimpy randomly picks a single word from a dictionary and presented with distortion and noise. The user needs to identify and enter the text correctly for authentication. The EzGimpy Captcha image is shown in Figure.2.

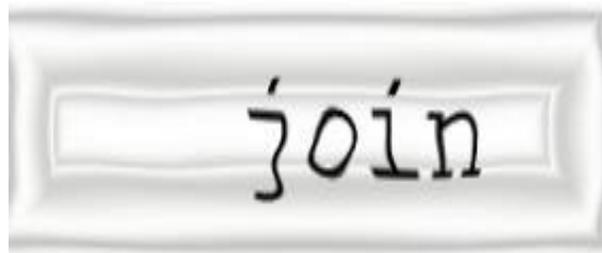


Figure.2. EzGimpy Captcha images

Baffle Text Captcha: Baffle Text Captcha [8] is a different type of text based Captcha introduced in California University, Berkeley. It is a customized version of Gimpy Captcha. It picks random alphabets or characters to form a pronounceable text. These words are garbage words not available in any dictionary. This is mainly to prevent the drawback of dictionary attacks. Figure.3. shows the Baffle Text Captcha image.



Figure.3. Baffle Text Captcha

Pessimist print Captcha: In the Pessimist print Captcha [9], a degraded low quality image consisting of characters was displayed to the users. The low quality images can only be read by humans and not by any OCR softwares. Therefore humans only can pass this Captcha test and it is very difficult for the bots. Figure.4. shows the Pessimist print Captcha image.



Figure.4. Pessimist print Captcha image

Handwritten Captcha: Handwritten Captcha [10], is another human interactive proof which challenges the bots with a handwritten image. An image of handwritten word in a noisy background is presented to the user. The user has to identify the characters for authentication. This will be a tough task for the computer bots, whereas it is easy for humans to recognize the handwritten words. Figure.5. shows the Handwritten Captcha image.

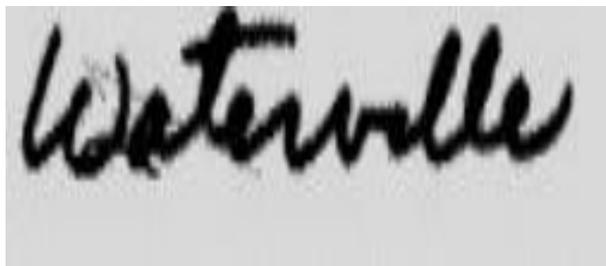


Figure.5. Handwritten Captcha image

ReCaptcha: ReCaptcha was introduced into the Google services, to prevent spam and also used to identify the deciphered words which are not recognized by the scanners. Out of the two words, one will be taken from an old book have to be digitized. The users need to identify both the two words for authentication. Figure.6. shows the Google ReCaptcha image.

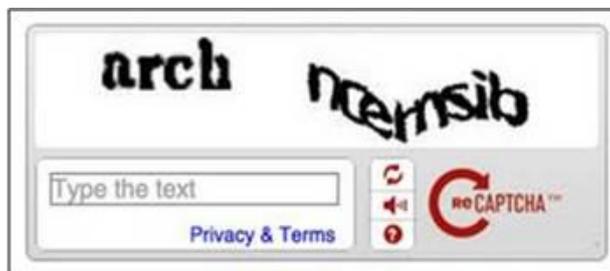


Figure.6. Google ReCaptcha image

MSN Captcha: MSN Captcha [11] is designed exclusively for Microsoft services. They are popularly known as MSN Passport Captcha. It is designed with a minimum of eight characters and digits, presented in dark blue foreground and grey background. Warping is also used to distort the characters. The ripple effect in the Captcha makes it very difficult for recognition by bots. The figure.7. shows the MSN Captcha images.

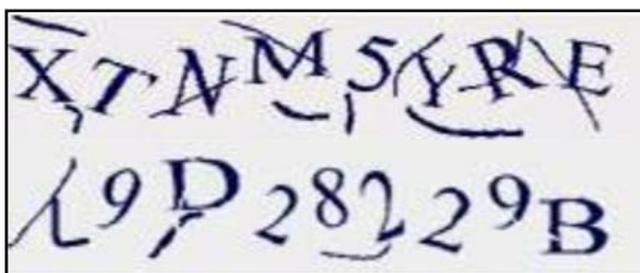


Figure.7. MSN Captcha images

3D Captcha: Human's visual structure is capable to identify 3D objects as well 3D movies. This will be a tough task for the computer bots to recognise 3D characters and objects. Hence a novel 3D Captcha [12] has been engaged with different orientations. This enhances the protection of websites from computer bots. In addition twisting and noise has also been introduced to increase the security. The Figure.8. shows the 3D Captcha image.



Figure.8. 3D Captcha image.

Captcha in popular websites: Based on the different concepts, a large number of popular websites engaged the text based Captcha methods to the security of their websites. The Captcha methods have been implemented with different modification and improvement. This includes lines, waving, overlapping of characters, color characters, noise in the background and degrading the image quality etc. The Figure 9 shows the Captcha images [13] employed in popular websites.

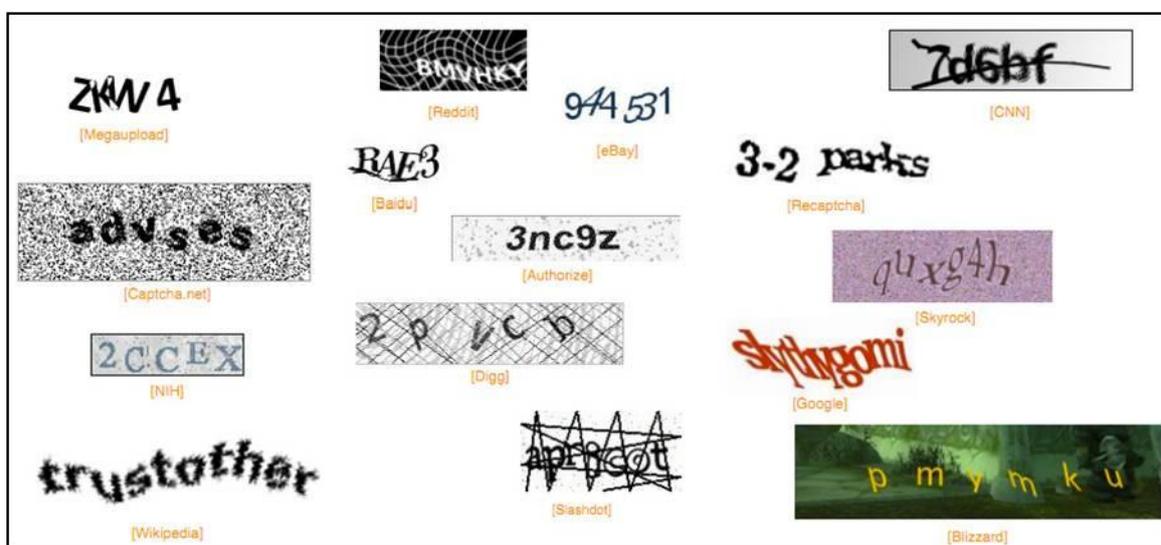


Figure 9. Captcha in Popular websites

Strength and Limitations of Text based Captcha methods: Every Captcha method has its own advantages, strength and limitations [14]. The commonly accepted and most widely used text based Captcha methods has so many advantages but suffers with some disadvantages also.

Advantages and Strength:

- Text based Captcha methods are user friendly to all section of users.
- It is easy for generation and validation
- Text based Captcha is simplest among all other types of Captcha
- It may be designed in every language
- It is very easy to understand and easy to answer for all humans

Disadvantages and Limitations:

- More distortion, different orientation and noise make it difficult for humans also.
- Text based Captcha can be cracked by using OCR softwares.
- It is suffered by relay, random guess and dictionary attacks.

V. CONCLUSIONS

The primary concerns of a Captcha design are usability and security. Every Captcha method should be easy for humans to solve and hard for bots to solve. Also it should be easy for generation and evaluation. If usability is increased, security will become a concern. If security is increased, usability will become a concern. Hence a perfect compromise between usability and security is essential in Captcha design. The text based Captcha methods are more user friendly hence better usability is achieved. If the security is further improved then it should be free from any kind of bot attacks.

REFERENCES

- [1] Luis von Ahn et al, 'CAPTCHA: Using Hard AI Problems for Security', in: Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques, Poland, 2003
- [2] M. Tariq Banday & N. A. Shah, 'A Study of Captchas for Securing Web Services,' International Journal of Secure Digital Information Age, 2009, Vol. 1. No.2.,
- [3] Henry S.Baird & kris popat, ' Human interactive proof and document Image analysis', in: International workshop on document analysis system, 2002,
- [4] Luis von Ahn, 'Labeling Images with a Computer Game', in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2004.
- [5] Gao et al , 'An audio CAPTCHA to distinguish humans from computers', in: 3rd International Symposium on Electronic Commerce and Security, pp 265–269, 2010.
- [6] Kiranjot Kaur & Sunny Behal, 'Captcha and Its Techniques: A Review', International Journal of Computer Science and Information Technologies, 2014, vol. 5, No.5.
- [7] Jeff Yan, Ahmad Salah & El Ahmad, 'CAPTCHA Security-A Case Study', IEEE Security & Privacy, 2009, volume: 7, Issue: 4.
- [8] Monica Chew & Henry S. Baird, ' Baffle Text: a Human Interactive Proof', in: Proceedings of the SPIE/IS&T Document Recognition & Retrieval Conference, 2003.
- [9] Coates, A.L, S. Bired & J. Fateman, 'Pessimal Print: A Reverse Turing Test', International Journal on Document Analysis and Recognition, 2003, vol.5, no.2, pp 158-163.
- [10] Amalia rusu , 'Handwritten Captcha using the difference in the abilities of human and machines in reading handwritten words', in : Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition, IEEE, 2004.
- [11] Baljit Singh Saini & Anju Bala, 'A Review of Bot Protection using CAPTCHA for Web Security' , Journal of Computer Engineering, 2013, volume 8, Issue 6 , PP 36-42
- [12] M. Imsamai &S.Phimoltares, '3D CAPTCHA: A Next Generation of the CAPTCHA', in: IEEE International Conference on Information Science and Applications, Seoul, 2010, PP 1–8.
- [13] Elie Bursztein, Matthieu Martin and John C. Mitchell, 'Text-based CAPTCHA Strengths and Weaknesses', in: Proceedings of the 18th ACM conference on Computer and communications security, October 2011, PP 125–138.
- [14] Rizwan Ur Rahman, 'Survey On Captcha Systems', Journal of Global Research in Computer Science, June 2012, Vol. 3, No. 6.