

Captcha and Its Techniques: A Review

Kiranjot Kaur

(M.Tech Student)

Department of Computer Science Engineering
SBS state technical campus
Ferozepur, Punjab

Sunny Behal

(Assistant Professor)

Department of Computer Science Engineering
SBS state technical campus
Ferozepur, Punjab

Abstract – Captcha (Completely Automated public Turing test to tell Computers and Humans Apart) system is used to verify whether a user is human or computer program. It's also known as human Interactive Proof (HIP) and based upon Artificial Intelligence. Captcha is a program that protects websites from web-bots by generating tests that computer cannot pass but human can pass. This paper comprise introduction of captcha, various techniques, application of captcha and drawbacks of captcha.

Index Terms – Captcha, Application, Issues and Drawback of captcha.

I. INTRODUCTION

Web services are major application provided by internet. As the use of web services are increase greater the chances of malicious robots attack on it. Captcha (Completely Automated Public Turing Test to Tell Computers and Humans Apart) system prevent various websites or used on web to avoid various bots to attack the network resources. A good quality captcha system must have the following characters: (1) Content can be easy to understand by human. (2) Quick and less time consuming. (3) Suitable for all type of bots abuse [1]. A high quality CAPTCHAs must be strong and easily usable [2].

Brief history of CAPTCHAs

Some theoretical methods that telling human and computer apart mentioned by Moni Naor in 1996. CAPTCHS test was run in the Alta-Vista web-server engine (1997) first time. In September 2000, Udi Manber of yahoo reported this 'chat room problem' [3]. Nicholas J. Hopper of Carnegie Mellon University, Manuel Blum and John, Luis von Ahn Langford of IBM found the term in 2000 [4]. The rest of the paper is organized as section 2 types of captcha. Section 3 application of captcha. Section 4 Issues in captcha. Section 5 drawback of captcha. Section 6 advantages of captcha.

II. TYPES OF CAPTCHA

1) *Text-Based CAPTCHAs*: This type of CAPTCHA is very easy to implement. Text-based Captcha is representing in distortion form that contains case-insensitive letters and digits. Text-based CAPTCHA is deployed in famous websites, examples Yahoo, Hotmail, Gmail, YouTube, PayPal etc [6]. Gimpy, Ez-Gimpy, Baffle-Text and MSN-Captcha are the types of Text-based Captcha. Figure 1 represents the Text-based captcha.



Figure 1: An Example of Gimpy Captcha



Figure 2: An Example of Ez-Gimpy Captcha

Gimpy CAPTCHA is based on optical character recognition (OCR). Gimpy select several words from dictionary and displays them on screen. The words that are displays in an image are corrupted and disfigure or misshape. Users should enter the words that are given in the image to attain access to the service.

Ez-Gimpy is the version of Gimpy Captcha that is developed by Henry Baird. In case of Ez-Gimpy, a single word is select from a dictionary and then misshaping is applied. The task is user identify the misshaped text correctly. It is not a good implementation and already defeated by OCRs. Ez-Gimpy was broken by Mori et al [15].

Baffle-Text is designed at California University at Berkeley by Henry Baird [5]. It's a modified version of Gimpy. In Case of Baffle-text; Figure 3 a random alphabets or characters are picked to form a pronounceable text. Then the user is challenged to enter the right word.

MSN-Captcha digits and 8 characters (upper case) are used. The foreground color is sets dark blue and background color sets grey. To produce the ripple effect and to distort the characters warping is used. MSN-Captcha is broken by yan [14]



Figure 3: An Example of Baffle-Text Captcha

2) *Image-Based CAPTCHAs*: Image-based CAPTCHAs are tests in which the users have to choose those images that have some similar properties [7]. Some types of Image-based CAPTCHAs are Pix and Bongo Captcha



Figure 4: An Example of Pix-Captcha

Pix CAPTCHA uses a large database of photographic and animated images of daily objects (a baby, a chair, a dog, a bird, etc); Figure 4 represents various samples. Set of images present in the Captcha system shown to user, all related with the same concept or object .The user must then enter the concept or object to which all the images belong [7]. For example – what is the similar feature among the following 3 pictures that shows in Figure 4 =” BABY”

Bongo CAPTCHAs (Figure: 5) is developed by Mikhail M. Bongard. In bongo visual pattern recognition problem is give to user for solving. Bongo contain two series of blocks, the left block series and the right block series. The blocks in the right series differ from those in the left, and the user should identify those properties that groups the two block series apart ; Figure 5 shows the Bongo Captcha [8].

3) *Video-Based CAPTCHAs* (Figure: 6): A video taken from public database that have three words that describe that video [9]. As the video plays words may submit, i.e. the user does not have to wait for the video to finish before submitting their three words. The user’s tag has to match to automatically produce ground truth tags then only the test is passed [10].

4) *Audio-Based CAPTCHAs*: Audio-Based CAPTCHAs are works upon the sound-based systems. These CAPTCHAs are produced for those who cannot see clearly. It has downloadable audio-clips [7]. In this user should listens and then submit the word. The first system which was named ECO was put into operation at City University in Hong Kong by Nancy Chan [8].

5) *Puzzle CAPTCHA*: In puzzle CAPTCHA (Figure: 5) a picture is divided into segments. A user is proposed to combine these segments to form a complete picture [13].

III. APPLICATION OF CAPTCHA

CAPTCHA have several applications for practical security, including

A. Protecting Website Registration

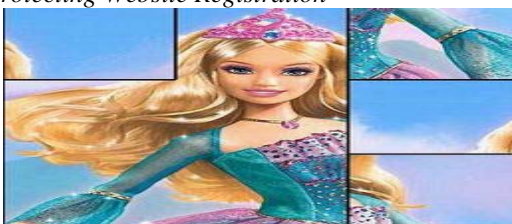


Figure 5: An Example of Puzzle-based Captcha

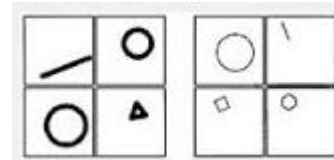


Figure 6: An Example of Bongo Captcha



Figure 7: An Example of Video-based Captcha

Many companies give free email services but generally suffered a specific type of attacks. To solve this problem CAPTCHAs to be used and it should made sure that only humans obtain free accounts and CAPTCHA is used to protect all free services of websites [1, 16].

B. Protecting Email Addresses from Scrapers

CAPTCHAs give full proof mechanism to hide your email address from scrapers. User is required to solve a CAPTCHA before showing email address [1, 16].

C. Online Polls

In online polling voters IP addresses were recorded so that single user should not vote more than once. A student of Carnegie Mellon found a way to stuffs the ballots using attack that voted for CMU many of time .So score of CMUs increase rapidly. Online polling should be protected with a CAPTCHA [1, 16].

D. Preventing Comment Spam in Blogs

Comments on a blog can be entered by humans only. It is not required to make users sign up before they enter a comment [16, 17].

E. Preventing Dictionary Attacks

To prevent dictionary attacks in password systems. After various unsuccessful login CAPTCHA is provided to user for login [16, 17].

F. Worms and Spam

CAPTCHAs also provide solution against email worms and spam [16, 17].

G. Search Engine Bots

The web pages should be left un indexed so that they cannot be easily founded .An html tag hinders search engine bots from reading web pages. But tag doesn’t guarantee, it only serves to say “no bots” please .Search engine bots belong to big company so do not see such pages. But to have complete guarantee that bots do not enter a web site CAPTCHAs needed [16, 17].

IV. ISSUES IN CAPTCHA

VI. Security Issue

There are two ways for breaking CAPTCHA challenge segmentation and character recognition. To divide of an image Content Based Image Retrieval (CBIR) methods are used. Success to a break a CAPTCHA technique depends upon the accuracy achieved in the segmentation process. Segmentation process chunks the image into letters and

passes chunks to the character recognition stage which attempts to map each chunk to a definite letter or digit [19]. After this process the techniques apply for recognition individual character at high speed. The complexity in segmentation depends upon the algorithms complexity used to generate the CAPTCHA test. It involves use of edge detection to divide an image into parts. It also detects outline of objects in an image by detecting jumps in its image intensity function. In simple pixels of image are set to white if their intensity is more than certain threshold value otherwise they are set to black. In some cases CAPTCHA produce images where adjoining characters have particular colors. Some use circles and lines as noise. Segmenting of overlapping characters is not easy [11].

B. Securing CAPTCHA against attacks

Many methods are used to make text-based CAPTCHAs difficult to break are font choice of letters, tricks, color model, noise, overlapping, draw lines, degradation and distortion [12]. Some random object like circles, lines, arcs and a complex background are used to secure CAPTCHAs against dictionary abuse [11].

C. Usability Issue

The use of colored background confuses the people and this reduces the scheme's usability. Low response time, Total accuracy and low perceived difficulty are required to make CAPTCHAs more useful [11]. Five quality components Efficiency, Learnability, Errors, Memorability, and Satisfaction define usability according to Jakob Nielsen. The CAPTCHAs nature ensures that the following usability criteria are applicable to address error, efficiency and satisfaction:

- *Accuracy*: how accurately can a user pass a captcha challenge? For example, to pass the test, how many times the user has to try?
- *Response time*: how much time taken by a user to pass the test?
- *Perceived difficulty/satisfaction of using a scheme*. How difficult to use do people perceive a CAPTCHA is? Are users satisfied and would users be willing to use this scheme? This group of criteria can be used for (quantitatively) measuring the usability of CAPTCHAs [2].

V. DRAWBACKS OF DIFFERENT TYPES OF CAPTCHA.

We describe various drawbacks of different types of CAPTCHAs [13]:

Text-based images:

- 1) In this type of CAPTCHAs, users have faced some problems to enter the correct text or characters or letter. Following are the some reasons that confuse the users to identify the correct text.
 - i. Use of various lines.
 - ii. Use of various shapes.
 - iii. Use of Multiple fonts.
 - iv. Font size variation.
 - v. Use of Blurred Letters

- 2) Text-based CAPTCHAs can be easily broken by OCR techniques (example: Content based image retrieval).

- 3) The peoples that have low visibility power cannot easily pass the test.

Image-based Captcha:

- 1) The People that have color blindness face many problems.
- 2) It can be identified by Random guessing attack or Pictionary-based attack [18].
- 3) Mislabeling, mis-spelling of images.
- 4) Large image databases or servers are required by these schemes to make the CAPTCHA. The times to make, grade and display the Captcha increases the load on the server

Video-based Captcha:

- 1) The size of files is large, so problem face by users to download video and pass the captcha test.
- 2) Speed of video.

Audio-based Captcha:

- 1) System available in the English so end user should have a comprehensive English Vocabulary.
- 2) Similar sound characters.
- 3) Not working for dumb people or people that have low listening power.

Puzzle-based Captcha:

- 1) Time consuming.
- 2) User cannot identify the puzzle easily.

VI. ADVANTAGES OF DIFFERENT TYPES OF CAPTCHA.

Text-based images:

- 1) Text-based captcha is simple to implement so it's mostly used in websites.
- 2) Battle Text-based captcha is used to defeat dictionary attacks.
- 3) Re-captcha Text-based captcha uses new dictionary words that cannot read using optical character recognition

Image-based Captcha:

- 1) Over the text-based captcha it increases the security.
- 2) Simple click based system so no need of typing.
- 3) Using Image-based captcha pattern recognition of image is difficult AI program.

Video-based Captcha:

- 1) It cannot break using Optical Character Recognition (OCR).
- 2) It cannot effect by laundry attacks.
- 3) In some cases it provides greater security than Text-based captcha and Image based captcha.

Audio-based Captcha:

- 1) It is used for people that have visual impairment.
- 2) Friendly to peoples.

Puzzle-based Captcha:

- 1) It seems like a fun.
- 2) It helps the user to monitor their brain.
- 3) It's like a game so user can more interact with this captcha system.

PERFORMANCE EVALUATION

[20] We can measure the performance of captcha using two factors. These two factors are security and usability. Using security and usability table 1 shows comparison between various captcha system
Image-based and Text-based captcha is secure and easily usable but not usable for vision impairs users. The only captcha that usable for blind users is audio-based captcha. Video-based and Audio-based captcha is difficult to use. Puzzle based captcha may or may not be easy and provide average security.

Table 1: Comparison of different captcha

Types of Captcha	Security	Usability	
		Usable for visible impair people	Easy or difficult to use
Text-based Captcha	Good	No	Average
Image-based Captcha	Good	No	Easy
Video-based Captcha	Good	No	Difficult
Audio-based Captcha	Good	Yes	Difficult
Puzzle-based Captcha	Average	No	Difficult/Easy

CONCLUSION

This paper, includes the different kinds of CAPTCHA have developed till now. In this paper on the CAPTCHA a brief study has been done, CAPTCHAs types and the applications of CAPTCHAs and drawbacks of different CAPTCHA based on Images, Text, Audio, Puzzles and Video. In future, main focus will be on to provide the CAPTCHA that easily access by the user and provide high quality of security that preventing the system from BOT abuse.

ACKNOWLEDGMENT

I am short of words to express our thankfulness to all the distinguished people who during my work gave me their unflinching support and valuable guidance, which helped me to make my endeavour a success. It is my pleasure to acknowledge Mr. Sunny Behal, Assistant Professor, Department of Computer Science & Engineering, SBS State Technical Campus, Ferozepur (Punjab), India, with deep sense of gratitude for their valuable time and help; they devoted on me and enlightened me with lot of knowledge for successful start and proper continuation of this project.

REFERENCES

- [1] Prof. Yogdhar Pandey, "Evaluating the Usability and Security of a Spelling Based Captcha System," *International Journal of Computer Science and Information Technologies(IJCSIT)*, Vol. 5 (3) , 2014, 4728-4731.
- [2] Ahmad Salah El Ahmad and Jeff Yan, "Usability of CAPTCHAs Or usability issues in CAPTCHA design," *Symposium On Usable Privacy and Security (SOUPS)*, July 23-25,2008, Pittsburgh, PA, USA.
- [3] Arturo Ribagorda and Carlos Javier Hernandez-Castro, "Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study," *Elsevier*, 2009, 141-157.
- [4] <http://en.wikipedia.org/wiki/CAPTCHA>.
- [5] H. S. Baird and M. Chew, "BaffleText: a Human Interactive Proof," *Proceedings of the 10th SPIE/IS&T Document Recognition and Retrieval Conference*, Santa Clara, CA, 305-316, 2003.
- [6] Khushpreet Kaur and Clark Pope, "Is It Human or Computer? Defending E-Commerce with Captcha," *IEEE Computer Society*, IT Pro March -April 2005.
- [7] www.slideshare.net/kunalkiit/seminar-report-on-captcha.
- [8] Anju Bala and Baljit Singh Saini, "A Review of Bot Protection using CAPTCHA for Web Security," (*IOSR-JCE*) *IOSR Journal of Computer Engineering*, Volume 8, Issue 6 (Jan. - Feb. 2013), 36-42.
- [9] Kurt A. Kluever, "Evaluating the Usability and Security of a Video CAPTCHA," *Master's thesis, Rochester Institute of Technology, Rochester*, New York, August 2008.
- [10] H. Kwak, M. chew, P. Rodriguez, S. Moon and Y.-Y. Ahn, "I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System," In Proc. IMC 2007, ACM Press, 1-14.
- [11] Ahmad El Ahmad , Jeff Yan and Wai-Yin Ng , "CAPTCHA Design Color, , and Security," *IEEE Computer Society*, 1089-7801, March – April 2012 IEEE.
- [12] D. Naccache, P. Gutmann and C. C. Palme, "CAPTCHAs: Humans vs. Bots", *IEEE Security & Privacy*, 68-70, 2006.
- [13] Preet Pal and Ved Prakash Singh, "Survey of Different Types of CAPTCHA," / (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, Vol. 5 (2) , 2014, 2242-2245.
- [14] Ahmad Salah El Ahmad and Jeff Yan, "A low -cost attack on a Microsoft captcha," in 15th ACM conference on computer and communications security, oct-2008, 543-554.
- [15] J.malik and G.mori, "Recognizing objects in adversarial clutter: Breaking a visual Captcha," *Conference on Computer Vision and Pattern Recognition*, vol-1,134-141, June 2003.
- [16] <http://www.captcha.net>.
- [17] Nitisha Payal, Nidhi Chaudhary, Parma Nand Astya, "JigCAPTCHA: An Advanced Image-Based CAPTCHA Integrated with Jigsaw Piece Puzzle using AJAX," *International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307*, Volume-2, Issue-5, November 2012.
- [18] Ashish Jain, Aditya Raj, Abhimanyu Jain and Tushar Pahwa1, "Picture CAPTCHAs with Sequencing: Their Types and Analysis," *International Journal of Digital Society (IJDS)*, Vol. 1, Issue 3, September 2010.
- [19] N. A. Shah, M. Tariq Bandy, "A Study of CAPTCHAs for Securing Web Services," *IJSDIA International Journal of Secure Digital Information Age*, Volume. 1, December 2009.
- [20] Rose Mary Abraham, Kumary R Soumya and Swathi K V, "A Survey on Different CAPTCHA Techniques," *International Journal of Advances in Computer Science and Technology (IJACST)*, Vol. 3, No.2, February 2014