CAPTCHA: A SECURITY MEASURE AGAINST SPAM ATTACKS

Pawar S.E[1], Bauskar Makarand M [2]

[1]H.O.D & Assoc. Prof., [2]Student, Information Technology, AVCOE, Maharashtra, India
suvrna.pawar@gmail.com, mbauskar@gmail.com

Abstract

A CAPTCHA is challenge response test used to ensure that the response is generated by humans. CAPTCHA test are administrator by machines to differentiate between humans and machine. Because of this reason CAPTCHAs are also known as the Reverse Turing Test as contrast to Turing Test which is administrated by humans.

CAPCHA is used as a simple puzzle, which restricts various automated programs (also known as internet-bots) to sign-up e-mail accounts, cracking passwords, spam sending etc. A common type of CAPTCHA requires user to recognize the letters from a distorted image, since normal human can easily recognize the CAPTCHA, while that particular text cannot be recognized by bot. In short CAPTCHA program challenges the automated program, which trying to access private data. So, CAPTCHA helps in preventing the access of personal mail accounts by some unauthorized automated spamming programs.

Index Terms: *CAPTCHA, Security, Spam Attacks, Reverse Turing Test.*

---***---

1. INTRODUCTION

Turing test was introduced by Alan Turing, the goal of Turing test was to serve as measure of the progress for artificial intelligence. If a computer passes the Turing Test then the computer is said to be intelligent. CAPTCHA uses the concept of Turing Test, in which CAPTCHA system acts as judge.

CAPTCHA is abbreviation which stands for Completely Automated Public Turing-test to tell Computer and Humans Apart. The Term CAPTCHA is introduced in 2000 by Luis von Ahn and his colleagues. A Turing test [2], named after famous computer scientist Alan Turing, is a method to differentiate a human from a computer. CAPTCHA employ the Reverse Turing Test as in CAPTCHA system administrate the test while in Turing Test normally human administrate the test.

Now days, websites became the identity of many business. Many companies have their own site for business. These company websites offers services for their users for example Gmail, yahoo, etc. In order to use these services user has to register on the website, so many people exploit such free services by duplicate registration. Such duplicate registration can be done by internet bots [1], it's a simple computer program that can automatically register on websites or such bots can also be use to put comments on the website. As these bots are computer programs it can it can register itself on the site hundreds or thousands of time without fail if proper security is not provided to the website.

To avoid such problem the term CAPTCHA is used. The CAPTCHA test is simple visual test or simple puzzle any human can crack this test or puzzle but automated program or internet bots will not be able to crack such challenges hence it will not able to gain access the services provided by the various websites. Most of CAPTCHA test include random codes in form of images, letter, alphabet and numbers that are overlapped over each other. These codes are easy for humans to understand, in order to gain the access user have to rewrite the code correctly. Example of CAPTCHA is shown in Fig 1.

Fig-1: CAPTCHA with distorted text

CAPTCHA has the following applications for practical security

- Online Polls. In Nov 1999, Slashdot.com posted an online poll asking which the best school in computer science was. Slashdot.com recorded the IP addresses of the voter to prevent single user from voting more than once. However, students of Carnegie Mellon University used a computer program which can vote for Carnegie Mellon University thousands of times. In such way the CMU's score stared to growing rapidly. Also MIT student also wrote their own voting program. MIT scored 21156 votes while Carnegie Mellon University scored 21032 and other schools score was less than 1000. This raised the question can the result of online poll be trusted? Not unless only human votes on such online polls. Here use of CAPTCHA can be effective as only human can pass the test, After solving the CAPTCHA we can decide whether it is human or automated program and based on result we let the user vote on the site.

- In order to provide better search result search engine bots browse the World Wide Web for web indexing. But some sites don't want their content to be indexed by search engines; They can use the html tags to prevent bots from reading the web pages. However, in order to be sure that no bots enter in web sites we can use CAPTCHA.

- There are many web site on world wide web that provides the services like mail services or message services for free. Internet bots can use such services to send the mail and message and as it is computer program it can send thousands or more mails or messages in a day. To avoid the misuse of such free services CAPTCHA can be use. Before Submission of mail or message we can ask user to solve CAPTCHA if the user successfully

2. BACKGROUND

Moni Naor (Naor, 1996) is the first person that mentioned some theoretical methods for telling apart computers from humans remotely. The First Know CAPTCHA test was used in Alta-Vista web search engine in early 1997. Alta-Vista faced the problem of automatic submission of URLs to their search engine. Andrei Broder at AltaVista and his colleagues developed a solution for this problem. Their method is to generate an image of printed text randomly so that machine character recognition systems (OCR) cannot read it but humans still can. [6]

In 2000, Yahoo describes their chat room problem to the Carnegie-Mellon University: bots were joining the chat room and pointing the chat room user to advertisement site. They wanted to prevent this bots from posting the advertisement in chat-rooms.

Professors Manual Blum, Luis A. von Ahn, and John Langford, from Carnegie-Mellon University, described some desirable properties of such a test:

1. The Test Challenges should be automatically generated and graded, that is, a computer program must be able to automatically generate the challenge and also mark the answer as qualifying or failed.
2. The test must be taken quickly and easily.
3. The test will accept virtually all human users with high reliability.
4. The Test will reject very few human users.
5. The Test will reject almost all machine users.

The Carnegie-Mellon University team developed a 'hard' GIMPY CAPTCHA which picked English words at random and rendered them as images of printed text under a wide variety of shape deformations and image distortions. A simplified version of GIMPY (EZ GIMPY, Fig. 2), using only one word-image at a time, was installed by Yahoo![7]

Gimpy CAPTCHA was built for yahoo to keep the bots out of chat rooms also to prevent script to obtain the email addresses. Gimpy CAPTCHA based on the human ability to read the distorted text.

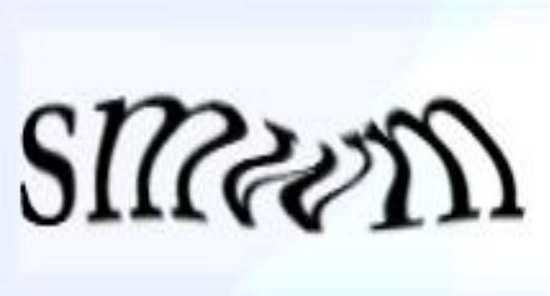


Fig -2: Gimpy CAPTCHA used by Yahoo!

3. EXISTING SYSTEM

Existing CAPTCHA algorithms can be grouped into three classes:

1. Text Based CAPTCHA
2. Graphics CAPTCHA
3. Audio CAPTCHA

Text-based CAPTCHAs are the most common and widely used now a days. The Text Based CAPTCHAs require the users to decipher text that has been visually distorted and rendered as an image. First CAPTCHA ever designed was AltaVista CAPTCHA. It used distortions of text that were known to reduce OCR accuracy [3]. GIMPY CAPTCHA which is similar to the AltaVista CAPTCHA [3,4], used English dictionary words. However, Mori and Malik showed that the GIMPY CAPTCHA can be broken and an attack rate of 92% was achieved against EZ-GIMPY [5] a variant of GIMPY. A major disadvantage of these early approaches of text based CAPTCHA was vulnerability to segmentation, where each character could be recognized in isolation. Windows MSN CAPTCHA introduced lines connecting individual characters though, it have high attack rates of 78% or more than that[3]. BaffleText's method of rendering a mottled black-and-white background and then performing various masking operations with overlapping text was more successful, being attacked in only 25% of the attempts [6].

Some CAPTCHA algorithms have taken an approach of using handwritten text images already known to fail optical character recognition rather than designing tests to be non-recognizable via OCR. In Such CAPTCHA system a database of text images gathered from handwritten mail addresses that couldn't be recognized by system automatically were used in such CAPTCHAs. When complete city names were used, humans were able to recognize the city name or hand written text 100% of the time but the computer success rate was about 9% [8]. reCAPTCHA was designed similarly using the text images scanned from book digitization projects [7]. In reCAPTCHA, users were presented with two text images and asked to enter both words. The known text served as the test while the currently unknown word's results were stored to help identify that word for future use. Researchers have found that the success attack rate for reCAPTCHA is in between 5% and 30% [9]. Examples of existing text CAPTCHAs are shown in Fig. 3.

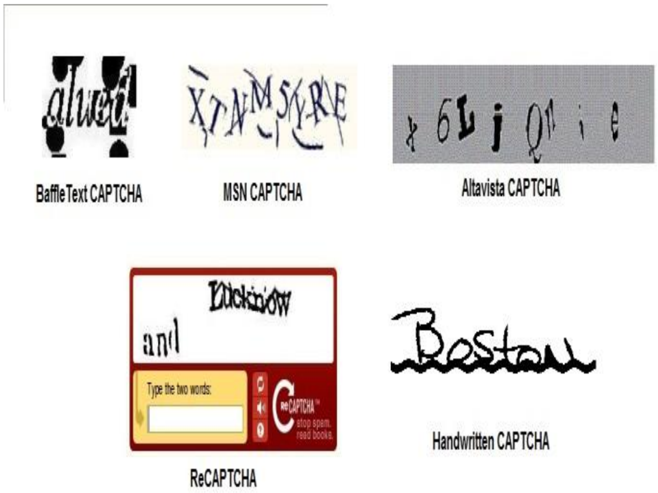


Fig -3: Example Of Text CAPTCHA

Several CAPTCHA applications utilize image classification or recognition tasks as part of their test [10]. One basic image-based CAPTCHA is ESP-PIX in which a collection of images are shown and the user has to select a description from a predefined list of categories [10]. Asirra is similar to KittenAuth and uses a closed database to source the images. These image-based CAPTCHAs demonstrate a common weakness a small number of possible solutions for which random guessing can have a high likelihood of success. MosaHIP requires dragging descriptors and dropping them on top of embedded images in a collage. Example of existing graphics based CAPTCHAs are shown in Fig. 4

These CAPTCHAs have high computer attack rates using a speech recognition approach. Specifically, the audio CAPTCHAs used by Digg and Google have a successful attack rate of about 70%. Example of audio CAPTCHA is shown in Fig 5.

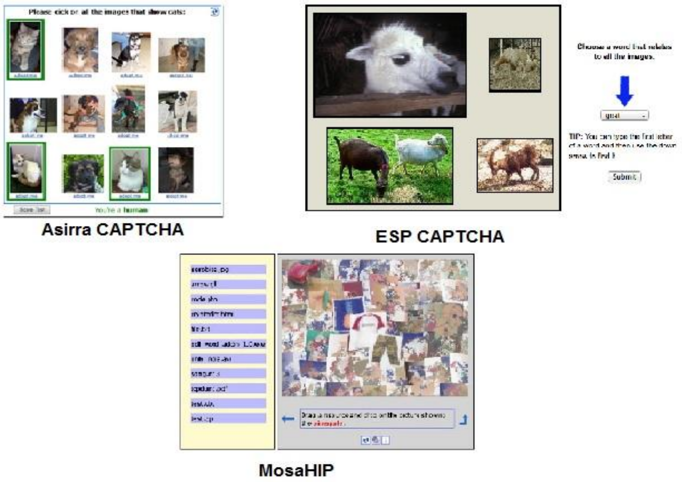
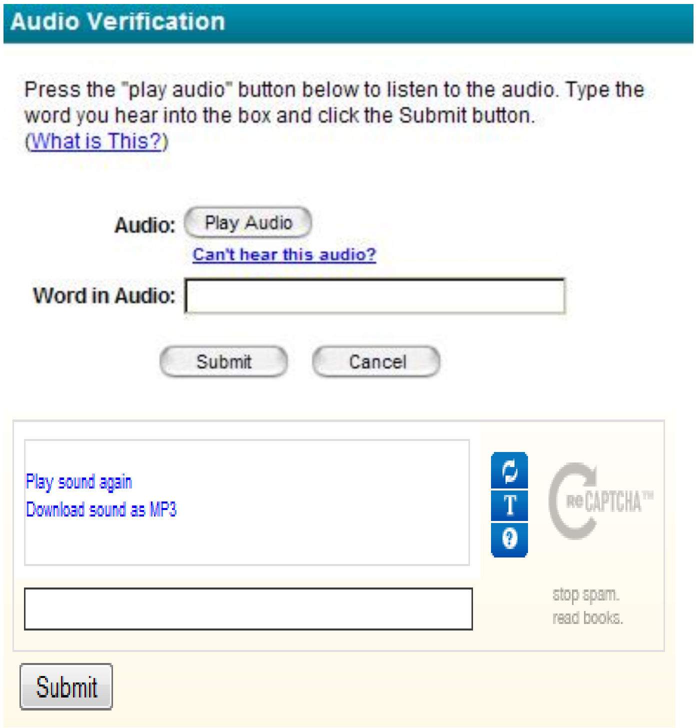

Please click on all the images that show cats:
Asirra CAPTCHA
Choose a word that relates to all the images.
Submit
ESP CAPTCHA
MosaHIP

Fig -4: Example of image based CAPTCHA



Audio Verification

Press the "play audio" button below to listen to the audio. Type the word you hear into the box and click the Submit button.
(What is This?)

Audio: Play Audio
Can't hear this audio?

Word in Audio:

Submit Cancel

Play sound again
Download sound as MP3

reCAPTCHA™
stop spam.
read books.

Submit

Fig -5: Example of Audio CAPTCHA

4. PROPOSED SYSTEM

In this paper, we've shown the problem of CAPTCHAs and offered enhancements to the interface wont to answer them that may facilitate build them a lot of usable. We have a tendency to attempt to explore different areas during which interface changes might improve visual and non-visual access, and take into account however the teachings we have a tendency to learn during this work might generalize on the far side the interfaces to CAPTCHAs.

Some of the CAPTCHAs face several issues, such as:

1. None are presently accessible to people who are each blind or have other vision issues and deaf ,
2. Automated techniques have become progressively effective in defeating them. A crucial direction for future work is addressing these issues.

Future work might explore however CAPTCHAs may be created that's easier for humans to resolve whereas still addressing the improved automatic techniques for defeating them. The CAPTCHAs explored in our study exhibited a good type of dimensions on that they varied, however nevertheless remained quite similar in style.

CONCLUSIONS

It is expected to open up new research in related field of network CAPTCHA. Putting forward creatively the essential framework and design principles and presenting a paradigm verification system won't solely facilitate to boost the usability that web users use the web Verification System however additionally improves the protection against malicious software package attacks. The analysis results are applied to broad areas of on-line user verification system, and have great value of theoretical analysis and practical application.

Many sites have millions of users and provide variety of services e.g. mail services, chat rooms that will always need access control that limits abuse. It is advisable to use visual or audio CAPTCHA to avoid the spamming attacks. It must be noted that user will pass through the system where internet bots and spam must be blocked by the CAPTCHA.

REFERENCES:

[1] The Turing Test, the Alan Turing Internet Scrapbook, 2002. The document is online at http://www.turing.org.uk/turing/scrapbook/test.html

[2] Luis von Ahn, Manuel Blum and John Langford, Telling Humans and Computers Apart Automatically: How Lazy Cryptographers do AI, In Communications of the ACM, 2004

[3] K.A. Kluever, Evaluating the usability and security of a video CAPTCHA, Master’s Thesis, Rochester Institute of Technology, 2008.

[4] H.S. Baird, K. Popat, Human interactive proofs and document image analysis, in: Document Analysis Systems, 2002, pp. 531–537.

[5] G. Mori, J. Malik, Recognizing objects in adversarial clutter: breaking a visual CAPTCHA, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, vol. 1, pp. 134–141.

[6] M. Chew, H.S. Baird, Baffletext: a human interactive proof, in: Proceedings of the Document Recognition and Retrieval, 2003, pp. 305–316.

[7] L. von Ahn, B. Maurer, C. McMillen, D. Abraham, M. Blum, ReCAPTCHA: humanbased character recognition via web security measures, Science 321 (2008) 1465–1468.

[8] A. Rusu, V. Govindaraju, Handwritten CAPTCHA: using the difference in the abilities of humans and machines in reading handwritten words, in: Proceedings of the 9th International Workshop on Frontiers in Handwriting Recognition, 2004, pp. 226–231.

[9] P. Baecher, N. Buscher, M. Fischlin, B. Milde, Breaking reCAPTCHA: a holistic approach via shape recognition, Future Challenges in Security and Privacy for Academia and Industry 354 (2011) 56–67.

[10] J. Yan, A.S. El Ahmad, Usability of CAPTCHAs or usability issues in CAPTCHA design, in: Proceedings of the 4th Symposium on Usable Privacy and Security, 2008, pp. 44–52.

[11] The official CAPTCHA site, http://www.captcha.net.

[12] ESP-pix, Carnegie Mellon University, http://www.captcha.net.

[13] Luis von Ahn, Manuel Blum, Nicholas J, Hopper and John Langford, The CAPTCHA Web Page: http://www.CAPTCHA.net, 2000

[14] L. von Ahn, M. Blum, and J. Langford. Telling humans and computers apart automatically. Communications of the ACM, 47(2):56–60, February 2004.

[14] Yan, J. and Salah El Ahmad, A. A Low-cost Attack on a Microsoft CAPTCHA, In CCS'08. Proceedings of the 15th ACM Conference on Computer and Communications Security, Alexandria, Virginia, USA, October 27-31, 2008

BIOGRAPHIES:

Associate Professor and Head of Information Technology Department, Amrutvahini College Of Engineering, Sangamner, Maharashtra

Bauskar Makarand M. completed his B.E. in information Technology from North Maharashtra University, in 2012. He is currently persuing his M.E. in Information Technology from University of Pune.