Survey on Captcha and its Techniques for BOT Protection

V.P remanand M. Tech. Student Department of Computer Science and Engineering, Manakula Vinayagar Institute of Technology, Pondicherry, India A. Meiappane, Ph.D. Associate Professor Department of Information Technology, Manakula Vinayagar Institute of Technology, Pondicherry, India. V. Arulalan M. Tech. Student Department of Computer Science and Engineering, Manakula Vinayagar Institute of Technology, Pondicherry, India.

ABSTRACT

The Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) is used to believe that a human is using the web service [1]. For the online services which are been provided in web services is been abused by automated bots over the internet. So CAPTCHA is used as a protection from these malicious programs for web security. But the bots are intelligent enough to break through these CAPTCHA. To increase the security, there are several CAPTCHA techniques has been proposed. In this paper various types of CAPTCHA techniques has been discussed and more secure techniques for the bot protection were discussed.

Keywords

CAPTCHA, Bot Protection, Automated computers.

1. INTRODUCTION

The CAPTCHA schemes have been proposed around the year 1999 to prevent the abuse of the hackers and the automated computers [2]. The CAPTCHA produces some simple grade tests which most of the humans can pass through, but the bots can't [3]. Thus the mostly used CAPTCHA are textual based type where the user must re-type the specified text given in the distorted manner. Then image CAPTCHA came into existence, where the images appear in the user interface and the user has to select the appropriate image from the given image in order to pass the CAPTCHA [4].

In both cases the user must recognize the text or image and then pass then CAPTCHA test, where it is pretty difficult for the bots. Then came some other types of CAPTCHA which is based on Non – Visual based, Feature based, and Riddle based automated test. Some exceptional computer programs which can break various CAPTCHA that have been proposed [5]. So the more effective CAPTCHA technique which gives more security over the bots and also easily solvable for humans are being considered as effective CAPTCHA technique. This paper is going to explain the various CAPTCHA techniques used from the beginning to till date which are used for the various internet services.

2. COMMON DEFINITIONS OF CAPTCHA

An effective CAPTCHA should differentiate the humans from bots and even an effective program can't break these tests, so the common definitions are:

- 1. Easily tackled by people
- 2. Easily created and asserted, but

3. The computer programs can't solve easily.

For the past 14 years many techniques over CAPTCHA have been proposed, where the security have been emerged day by day. An effective CAPTCHA must be a basic thing for the user to solve with involving a basic knowledge of the user so that most of the users passes these tests and the user's time is not wasted in solving these CAPTCHA. Since a user can pass through these CAPTCHA easily but it is intractable for the computer programs because the basic knowledge is been used and it is very costly to design the systems to break these CAPTCHA than using human resources.

The commonly used CAPTCHA are Visual CAPTCHA where the text (or) image is been muddled by changing the visual by the addition of some irregular fluctuations and distortion. Thus the challenge here is to design the muddling of visuals which is also easy for the users and they are not discouraged from attempting a CAPTCHA, yet making it too hard for the computer's vision algorithm to pass through. The major types of captcha are

- i. Visual based CAPTCHA
- ii. Non-Visual based CAPTCHA
- iii. Feature based CAPTCHA
- iv. Riddle based CAPTCHA

3. VISUAL BASED CAPTCHA 3.1. Text based CAPTCHA

A text based CAPTCHA is an interesting form of user interface where the point of recognition comes under reading based technique composed of English characters and Arabic numerals which are generated randomly in a distorted manner which is entered by the user in a text – box or rotating the texts in a random manner in a form of equation or a text with symbols and noise in small variations in fonts [4].

Thus we make the text based CAPTCHA is easy to solve by humans but difficult for the computer programs because of the distortion, noise, color aspects of the visuals. The distortion is been done vertically and horizontally thus the letter is been appeared twisted out of shape which becomes more difficult for the computer programs to identify.



Figure 1: A typical text based CAPTCHA

To improve security some more advancements in text CAPTCHA is been proposed one of them is handwritten CAPTCHA which are more secure than a machine printed CAPTCHA [6] which uses images of a handwritten random words which are been used to increase the complexity of the recognition of CAPTCHA by some advanced computer algorithms[7]. And many other text based CAPTCHA came into existence which are very easy to implement and also very effective.



Figure 2: Handwritten CAPTCHA

And now more advanced text based CAPTCHA techniques came to existence like reCAPTCHA which is developed by google and is a free service for the web pages provides more security for the bots. Here two letters are given one is an unknown word having combination of alpha numeric characters and another is a proper known word and this method eliminates the dictionary attacks.



Figure 3: reCAPTCHA

3.2 Image based CAPTCHA

Graphical CAPTCHA are the test challenges where the user must identify some images which have some similarity [8]. They are given in the form of visual puzzles so that the user must identify the image and type the text below (or) identify the text and click the image below. Other approach is that the user must find a common feature among some set of images.



that it is a second sec

Figure 4: Naming the Images



Figure 5: Identifying the oddity

Thus Graphical based CAPTCHA are of three types,

3.2.1 Naming the images

The images which have similarities will be given in a grid so that the user must identify the set of images and type the common word from the images (figure 4). In this figure 4 the common term is astronaut. But here the user may confuse with the images and the guessing may be of a similar synonym which may go wrong. So the CAPTCHA must be provided by hints to the user which makes the system a bit complicated.

3.2.2 Distinguishing the images

This type presents two sets of images to the user, each set contains three images describes same subject or different ones. The user must determine whether the images having same subject or not. For example three different images of table and other three images of a cat, but polysemy may be a problem here.

3.2.3 Identifying the oddity

Here a grid containing images of same subject, but only one image will be given different (odd image). The user must identify the odd image to pass the test (Figure 5). Here misspelling of words is not a problem, because the user is just going to click on the odd image. Polysemy may be a problem for here too, example a rat and a pointing device of a computer gives a same word 'mouse' which occurs in rare cases.

3.2.4 Identify the image

In this type a grid containing set of images, and name of a particular image is given above, the user have to identify the image name and should click on the identified image. The user should just click on the particular image given on top of the grid so it is simpler than other image CAPTCHA technique. But the coexistence of numerous conceivable implications for a word may be a problem because two images of denoting a same word may confuse the user which occurs in rare case.



Figure 6: Identify the image

4. NON – VISUAL BASED CAPTCHA

The audio CAPTCHA is an example for a non-visual based CAPTCHA which is based on sound systems which is been designed for the users who are unable to see (Physically challenged) [9]. In this type the CAPTCHA comes with an audio - clip consisting of a spoken word of a human which is a downloadable content. The user should play the audio content and type in the given text – box.

The audio - CAPTCHA differentiates the computer bots and humans by recognizing the spoken language. The CAPTCHA consists of chosen sequence of letters and digits which is randomly spelled by a sound clip. Then the sound clip is been provided by a distortion and then presented to the user to identify and enter the right number or word.



Figure 7: Audio Captcha

To improve the user interface of the existing audio CAPTCHA, user is allowed to control the playback of the clip. Example: Pause / Play, replay etc. Most of the audio

CAPTCHA consists of human voice with a finite vocabulary and noise are stronger CAPTCHA.

5. FEATURE BASED CAPTCHA

A video CAPTCHA is referred as feature based CAPTCHA, which is a newer technique but very rarely seen in internet [10, 4]. In this system 3 specific words are given in a running video format having common property with other words (say red colored words) so the user should identify the specified words from the moving words and type in given text box. One of the commercial system in the video – CAPTCHA is NuCAPTCHA a Canadian based organization for NuData security.



Figure 8: Video based CAPTCHA



Figure 10: NuCAPTCHA Moving letters

For a user to pass the CAPTCHA challenge, they must view a short video and enter the described tags for it. More than 60% of users say that a video CAPTCHA is "More Enjoyable" than other CAPTCHA methods because it is comfortable on human's eyes and more secure for bots.

6. RIDDLE BASED CAPTCHA

In a riddle based CAPTCHA an image is divided into lumps so that the user should be able to combine these lumps by drag and drop format and replicate the original image to pass the CAPTCHA test [11]. Generally the humans can perform this task well but the bots can't resulting in highly secured system. The main advantage is this system is a language independent so that any user can pass the test.



Figure 11: Riddle based CAPTCHA

7. CONCLUSION

In this paper we have discussed about various techniques of CAPTCHA existing till date. A brief description about visual based, non-visual based feature based, and riddle based CAPTCHA are given and explained clearly. This work can be used in various fields of security to avoid problems related to Artificial Intelligence. By using CAPTCHA as a graphical password we can ensure security better than other text based passwords which use hard mathematical cryptographic methods. Our future work will focus on the conversion of these CAPTCHA techniques into a Graphical passwords and providing security for the hard artificial problems.

8. REFERENCES

- L. von Ahn, M. Blum, and J. Langford, "Telling Human and Computers Apart Automatically," in Communications of the ACM, vol. 47, February 2004, no. 2, pp. 57-60. DOI:10.1145/966389.966390.
- [2] H.S. Baird and K. Popat, "Human Interactive Proofs and DocumentImage Analysis," Proceedings of the 5th IAPR

International Workshop on Document Analysis Systems, Princeton, LNCS 2423, 2002, pp. 507- 518.

- [3] Baljit Singh Saini and AnjuBala "A Review of Bot Protection using CAPTCHA for Web Security," IOSR Journal of Computer Engineering, 2013, pp. 36-42, 2013.
- [4] Chandavale, A.A., Sapkal A.M and Jalnekar, R.M."Algorithm to break visual CAPTCHA" Emerging Trends in Engineering and Technology (ICETET), 2009 2nd International Conference, 10.1109/ICETET.2009.24
- [5] Ahn, L. von, Blum, M., Hopper, N. J., & Langford, J., (2003), "CAPTCHA: Using hard AI problems for security", Proceedings of Eurocrypt 2003.
- [6] Mukta Rao and Nipur Singh "Random Handwritten CAPTCHA: Web Security with a Difference" I.J. Information Technology and Computer Science, 2012, 9, 53-58 Published Online August 2012 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijitcs.2012.09.07
- [7] Xiao Ling-Zi and Zhang Yi-Chun "A Case Study of Text-Based CAPTCHA Attacks" Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2012 International Conference 10.1109/CyberC.2012.28.
- [8] Monica Chew and J. D. Tygar "Image Recognition CAPTCHAs" In Proceedings of the 7th International Information Security Conference (ISC 2004), Springer, September 2004, pp. 268-279.
- [9] J.P. Bigham, and A.C. Cavender, Evaluating existing audio CAPTCHAs and an interface optimized for nonvisual use. In Proceedings of ACM CHI 2009 Conference on Human Factors in Computing Systems, 2009, pp. 1829-1838.
- [10] Shirali-Shahreza. M and Shirali-Shahreza. S "Motion CAPTCHA" Human System Interactions, 2008 Conference DOI:10.1109/HSI.2008.4581589.
- [11] NitishaPayal, Nidhi Chaudhary, Parma NandAstya "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.