Survey of Different Types of CAPTCHA

Ved Prakash Singh, Preet Pal

School of Computer Science, Lovely Professional University
Phagwara, Punjab

Abstract— **CAPTCHA is an acronym for Completely Automated Public Turning Test to tell Computers and Humans Apart [1]. CAPTCHA is basically used as a protection from these malicious programs like Bot. Now a day's for web security we are using different type of captcha. In this paper we describe all type of captcha and also describe their drawbacks all type of captcha. We also describe application of captcha. And review paper of different types of Captcha.**

*Index Terms—***CAPTCHA, BOT, Turing Test, Security, HIP OCR.**

I. INTRODUCTION

First time CAPTCHA was invented in 2000 at Carnegie Mellon University by John Langford, Nicholas J. Hooper and Luis Von Ahn [1]. CAPTCHA is an acronym for "Completely Automated Public Turning Test to tell Computers and Humans Apart" [2]. The progress of Internet, Web security has become an important issue. There are too many malicious threats across the Internet which may compromise your system in the absence of any secure application which provides protection against such threats. One such threat is the Bot. A Bot is a malicious program which has the capability to run automated tasks over the network and thus creating problem in the network [3]. CAPTCHA is one such shield which can be used as a protection from these malicious programs like Bot.

A CAPTCHA may come in various forms like text based or image based CAPTCHA. The Bot operation is similar to reverse "TURING TEST" (given by Alan Turing) [2] where the program acts like judge and the other person acts like user. If the user fails this test then he/she is considered to be a machine otherwise the user is considered to be an authentic user or a human being. CAPTCHA is a defensive system that acts as a tool to check web Bots from exploiting online services on the internet including free email providers, wikis, blogs etc. It is a HIP system that is widely used to secure the internet based applications. It is also called as a challenge response test which gives a challenge to the users, when the user gives correct answer he is considered as human otherwise a web bot. CAPTCHA is an authentication process based on challenge-response authentication. CAPTCHA provides a mechanism with the help of which a user's can protect themselves for spam and password decryption by taking a simple test. In this test a user will see either an image or a text which are normally distorted. The user is supposed to enter the pattern exactly as shown to him if the CAPTCH is based on text. If the CAPTCHA is based on image the user is supposed to enter the correct name of the image which correctly symbolizes the image. CAPTCHA is used where authenticated access is the primary concern. Various web services like Yahoo, Google, and Bing etc. use CAPTCHA to differentiate between an authenticated user and a malicious program. CAPTCHAs are also used in the sites which provide access to sensitive data, such credit card accounts and banks.

II. CATEGORIES OF CAPTCHA

CAPTHCAs means presenting a challenge response test to the users or humans. They are classified based on what is distorted that is whether characters, digits, or images. Some types of CAPTCHAs are given below:
1. CAPTCHAs based on text.
2. CAPTCHAs based on image.
3. CAPTCHAs based on audio.
4. CAPTCHAs based on video.
5. CAPTCHAs based on puzzle.

1) *CAPTCHAs based on text:* Text based CAPTCHAs is a very simple to implement. It is very effective and requires a large question bank. In Text based captcha the Number of classes of characters and digits are very small so the problem occur for user to identify the correct characters and digits. The text based captcha is possible to identify the character and digit through Optical character recognition (OCR) technique [2, 4]. In Text based CAPTCHAs simple asked questions like as based on arithmetic equation some example are given below:
A) What is three plus two (3+2=?).
B) What is six minus one (6-1=?).
C) Which of cabbage, apple and table is vegetable?
Example of text based CAPTCHA.
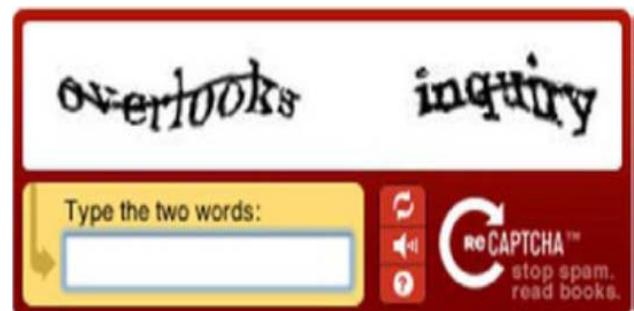
Fig 1: Text Based Captcha

2) *CAPTCHAs based on image:* Graphics-based CAPTCHAs are challenge-tests in which the users have to guess those images that have some similarity. For example: visual puzzles. In image based CAPTCHAs user is required to identity image. The advantage of image based

CAPTCHA is that pattern recognition is hard AI problem and therefore it is difficult to break this test using pattern recognition technique. Example of images based CAPTCHA are given below.
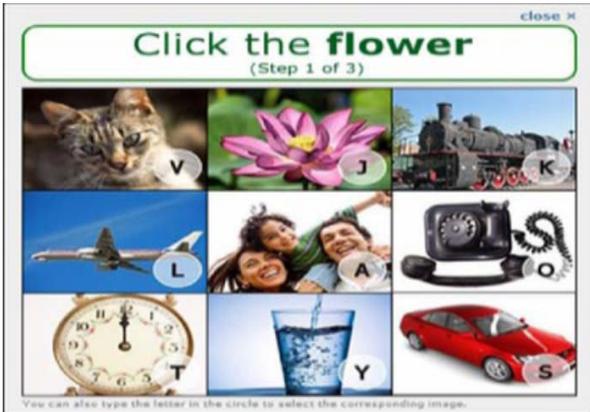
Fig 2 Images Based Captcha

3) CAPTCHA based on audio: Audio-based CAPTCHAs are based on the sound-based systems. These CAPTCHAs are developed for visually disabled users. It contains downloadable audio-clips. In this type of CAPTCHA, first the user listens and after that submits the spoken word [2]. The first sound-based system name ECO was implemented by the Nancy Chan a student from the City University in Hong Kong. The audio-based system is based on the difference in the ability between computer machines and humans in recognizing spoken language. The program chooses a sequence of digits and words randomly and renders the words and number digits into sound clips and distorts it. The distorted sound clip is then presented to the user to enter the right word or number. The user is asked to enter exactly the same words as spoken the audio clip [6]. Example of audio based CAPTCHA are given below.
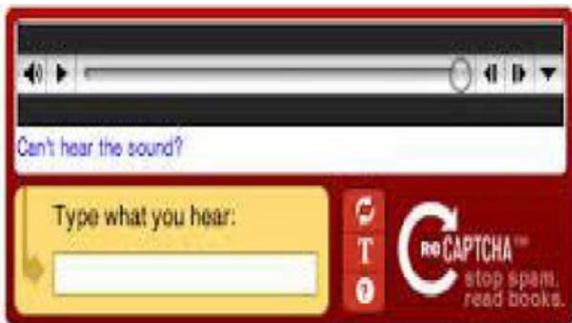

Fig 3 Audio Based Captcha

4) CAPTCHA based on video: Video CAPTCHA is a newer and less commonly seen CAPTCHA system. In video-based CAPTCHAs, three words (tags) are provided to the user which describes a video. The user's tag must match to a set of automatically generated ground truth tags then only the test is said to be passed. The term video CAPTCHA is used to any CAPTCHA that uses a video as its means to present information to a user [2, 6]. Although video CAPTCHA is limited, both commercial and

academic application do exist. Example of video based CAPTCHA are given below.

Fig 4 Video Based Captcha

5) CAPTCHA based on puzzle: Usually in puzzle based CAPTCHA a given picture is divided to chunks [2, 5]. A user is supposed to combine these chunks so as to form the complete picture same as the original one.

Example of puzzle based CAPTCHA are given below.
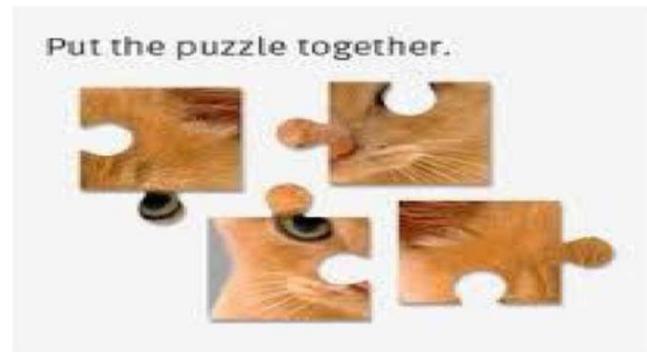
Fig 5 Puzzle Based Captcha

III. LITERATURE SURVEY

. I have studied the research papers related to the different types of captcha, as given below

Wei-Bin Lee Che-Wei Fan et al. [1] In their research paper they have mentioned a new mechanism for text based CAPTCHA in which a user can recognized the alphabets from a highly distorted image with the help of a tip. This mechanism has been designed in such a way that it is very hard for a computer program to recognize the pattern of text but will be easier for a human to interpret. In this mechanism the text given to the user is the mixture of upper case letters and lower case letters with high degree of distortion. With each text a tip is given so that a user can interpret the text based on that tip independent of the degree of distortion. A user is supposed to identify each character, either in lower or upper case, and to enter exactly the same stream of characters that has been shown..

Baljit Singh Saini et al. [2] In this paper, they examines CAPTCHAs and its working and literature review and they also describes classification of CAPTCHAs and its application areas and guidelines for generating a CAPTCHA. To implement the CAPTCHA there are three methods: Visual method, OCR (Optical Character Recognition) and non OCR visual method. In this paper

they are comparison types of CAPTCHAs. This paper gives a description for various existing CAPTCHA schemes with a literature survey and provides a description for working of CAPTCHA. This is also describes the classification of various CAPTCHA schemes. And finally, the various applications of CAPTCHAs and comparison between OCR and NON-OCR based CAPTCHAs are also presented. Today various daily activities like entertainment, education, e-commerce etc. are carried out by using the Internet. To perform such web activities users have to register regarding the websites. In registering websites, some intruders write malicious programs that waste the website resources by making automatic false enrolments that are called as bots. These false enrolments may harmfully affect the working of websites. So, it becomes necessary to differentiate between human users and Web bots (or computer programs) is known as CAPTCHA. CAPTCHA is based on identifying the distorted text, the color of image, object. This paper examines CAPTCHAs and its working and literature Review. This paper also provides classification of CAPTCHAs, its application areas and guidelines for generating a CAPTCHA. There are three basic properties that CAPTCHAs must satisfy.

1. CAPTCHAs should be easy for human user pass.
2. It should be flexible enough so that a tester machine easily generate and grade it.
3. Must be hard enough for a bot to pass.

Xlao Ling-Zi et al. [3] In this paper they describe about CAPTCHA which is based on text. It has been mostly used because of easily accessible on the internet. The CAPTCHA implementation is very risky without careful design .There are many attacks on text based CAPTCHA system. Most of result from poor security and careless design overlook. For this purpose they design good and well security to prevent from attacks.

Chen-Chiung Hsieh et al. [4] In this paper they proposed an innovative image-based CAPTCHA for distinguishing human and computer by embedding versatile characters in the images and in method who they proposed it makes the characters invisible by automated image analysis technologies like scale-invariant feature transform while human can easily distinguish the location of the embedded characters . Their designed mechanism was capable to elude such generous of attacks. For in experiments, 15 users were invited to test the system and the success rate is 86%. If wrong operations like clicking out of text boxes were excluded, the success rate reached 95%. Compare the average logging time with reCAPTCHA and hello CPTCHA, the proposed method is faster than these two methods by 32 seconds and 115 seconds, respectively. In this paper they proposed random number generation module is used to select image and verification characters for conducting man machine distinguishing. Then color, size, style, and angle variation are applied to the verification characters. Finally, the verification image is generated.

S.Benson Edwin Raj et al. [7] Denial of Service is a common threat to network security and is also considered to be an automated network attack. To prevent the system from such kind of attacks the CAPTCHA shield must be implemented in the system to identify the difference between the legitimate user and the fake one. These days OCR based CAPTCHAs are more vulnerable as there are too many correlation algorithm available in the Internet with the help of which these CAPTCHAs can easily be broken. There is a way to cope with it and that is to make the CAPTCHA text more distorted, but it has its own disadvantage as the user won't be comfortable to read such kind of CAPTCHAs. To overcome this problem a new CAPTCHA was introduced called picture based or image based or graphics based CAPTCHA. This CAPTCHA based on images has its own advantages as any malicious program can't perform any type of segmentation through edge detection and thresholding, shape matching and random guessing. In the security analysis process this mechanism has shown better results.

Aditya Raj et al. [8] CAPTCHAs based on OCR have been exploited and are thus insecure, while the CAPTCHAs based on non-OCR are safe to use as they employ natural skill of the person's eye to identify the picture. They incorporated the concept of Sequencing in Picture CAPTCHA to build a new concept known as SPC. SPC generation can be classified in to two types – inherent sequencing and non-inherent sequencing. The former does not require tags while the latter does. SPCs include two levels of security, the first one is the recognition of objects in pictures and the second one is to determine their logical sequence.

Rich Gossweller et al. [5] Based on identifying an image's upright orientation a new CAPTCHA mechanism was introduced. In this mechanism the complex contents of an image are analyzed which is generally a tough task for the machines to do but a human eye can do it without any hassle. The algorithm used in this mechanism chooses the images from the repository, usually from the search result, that can be automatically set upright easily. They then apply a mechanism which verifies that the remaining images comes with human recognizable upright orientation. The most unique feature of this technique is that it is language independent and does not require any text entry for the verification of the user. Its implementation is very rapid in nature and there is full supply of images. They have performed many experiments to test the rigidity of this mechanism.

JingSong Cui et al [9] In this research paper, a new CAPTCHA has been mentioned which is based on moving object identification and tracking problems. It is referred to biological motion vision model. Based on Edge Mutation an innovative Single-frame Zero-knowledge rule is also put forward to the CAPTCHA generation algorithm. Only after solving the moving object recognition problem successfully an attacker can access the test service system. These kind of animated CAPTCHA are not only able to check the attacks based on static OCR technology but also check the attacks against the moving object detection. Inside the research paper they have mentioned three kinds of programs: the non-visual programs, the visual programs based on OCR problems and visual programs based on non-OCR problems. It most widely used and applied program is based on the visual programs based on the visual programs based

on OCR problems with the advantage of implementation and operation. Whenever a user makes a request to the server, the server respond to the user with a picture containing a string of random characters and numbers. The user has to identify the sequence of characters in order to get access to the server resources.

IV. DRAWBACKS OF DIFFERENT TYPES OF CAPTCHA.

S.no	Different types of captcha	Drawbacks
1	Text based captcha	1. In text images, user has some problem to identify the correct text or characters. i. Multiple fonts. ii. Font size. iii. 3.Blurred Letters iv. Wave Motion. 2. It can be easily identified by OCR techniques.
2	Images based captcha	Some users face problem of image identification who have low vision or due to blurring of images.
3	Audio based captcha	1. It is available in English therefore end user must have a comprehensive English vocabulary. 2. Character that have similar sound.
4	Video based captcha	Due to large size of file, users face problem to download video and find correct captcha
5	Puzzle based captcha	The task is not easy for users because puzzle based captcha take more time to solve the puzzle and identify actual arrangement of puzzles.

V. APPLICATION OF ALL TYPES OF CAPTCHA

There are number of applications of CAPTCHA on the web which are defined as follows.

1) Registering the web forms: There are many sites on the Internet which provide free registration to avail their services. But they are susceptible to web bots. It may come into the form of scripts which can register thousands of email accounts on the internet, thus wasting the precious space of web.

2) Online polling sites: These sites takes user's response or feedback in the form of questionnaires. To ensure that only human makes the response they make use of CAPTCHA.

3) To avoid web crawling: If a site doesn't want to get indexed by a search engine then they can make use of CAPTCHA.

4) E-Ticketing.

5) Preventing Dictionary Attacks and E-mail spam.

CONCLUSION

In this paper, we have studied the different kinds of CAPTCHA have developed yet. A brief survey has been done on the CAPTCHA and list out the applications and drawbacks of different CAPTCHA based on Text, Images, Audio, Video and Puzzles. In near future, main focus will be on to develop CAPTCHA that provide the ease of access to the user and highest level of security by preventing the BOT attacks.

REFERENCES

[1] Wei-Bin Lee, Che-Wei Fan ,Kevin Ho, Chyi-Ren Dow , and "A CAPTCHA with Tips Related to Alphabets Upper or Lower Case," in Seventh International Conference on Broadband, Communication, Wireless Computing and Applications, 2012.

[2] Baljit Singh Saini and Anju Bala "A Review of Bot Protection using CAPTCHA for Web Security," IOSR Journal of Computer Engineering, 2013, pp. 36-42, 2013.

[3] Xiao Ling-Zi and ZHANG Yi-Chun "A Case Study of Text-Based CAPTCHA Attacks," in International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover, 2012.

[4] Chen-Chiung Hsieh and Zong-Yu Wu "Anti-SIFT Images Based CAPTCHA Using Versatile," IEEE, 2013.

[5] Rich Gossweiler, Maryam Kamvar and Shumeet Baluja "What's Up CAPTCHA?A CAPTCHA Based on Image Orientation" WWW 2009 MADRID!, pp. 841-850, 2009..

[6] Wenjun Zhang, "Zhang's CAPTCHA Architecture Based on Intelligent Interaction via RIA," Research Institute of Applied Computer Technology, IEEE, 2010.

[7] S. Benson Edwin Raj, Deepa Devassy and Jiji Jagannivas "A New Architecture for the Generation of Picture Based CAPTCHA," IEEE, pp. 67-71, 2011.

[8] Aditya Raj, Ashish Jain, Tushar Pahwa and Abhimanyu Jain "Picture CAPTCHAsWith Sequencing: Their Types and Analysis," International Journal of Digital Society, vol. 1, no. 3, pp. 208-220, 2010.

[9] JingTing Mei , JingSong Cui, Da Zhang, LiJing Wang, Xia Wang, Yang Peng, WuZhou Zhang "CAPTCHA Design Based on moving Object Recognition Problem," in International Conference on Cyber-Enabled Distributed Computing and Knowledge Discover, 2012