A Survey on Different Types of CAPTCHA

Pavankumar Kulluru
Asst. Professor
Dept. Computer Engineering,
G.H.R.C.E.M Pune,
India

Taslim Shaikh
Student
Dept. Computer Engineering,
G.H.R.C.E.M Pune,
India

Shreepad Bidwai
Student
Dept. Computer Engineering,
G.H.R.C.E.M Pune,
India

Aditya Mannarwar
Student
Dept. Computer Engineering,
G.H.R.C.E.M Pune,
India

Kapil Shamlani
Student
Dept. Computer Engineering,
G.H.R.C.E.M Pune,
India

ABSTRACT
The world today is filled in with enormous amount of innovation and technology, Internet is thus stated to be one of the most beneficial part awarded to the society ,but web security is the crucial issue to be considered to ensure this a concept called CAPTCHA is developed. CAPTCHA stand for Completely Automated Public Turing test to tell Computer and Human Apart this is a technique that was initialized in 1997 and was developed in 2000.

The term "CAPTCHA" was coined by Luis Von Ahn, Manuel Blum, Nicholas J. Hopper (all of Carnegie Mellon University, and John Langford (then of IBM).CAPTCHA is an automated Turing test that can help us to generate and test grades which human can easily understand and pass while bots cannot.In this paper we are analyzing and using the concept of Audio ReCAPTCHA**.**

Keywords
Captcha, Recaptcha, Audio,Digitization

1. INTRODUCTION
ReCAPTCHA is a service that is freely available which ensures protection of your website from spam and abuse. ReCAPTCHA makes a use of an advanced risk analysis engine and adaptive CAPTCHAs for the purpose of obtaining automated software from engaging in abusive and unauthorized activities on your site [1]. It does this by allowing a valid user to pass through it with ease.

The ReCAPTCHA is an open captcha service that serves for a purpose for digitizing books and daily paper, to buckler your site from spam and bots. The ReCAPTCHA service gives two words that are randomly drawn from the database of server. The first word is control word and another one is a questionable word or vice-versa. The user interprets the words and after that the framework expects that if user sorts the control word accurately, then the questionable word are right after this the word is send to outcome the digitization of projects. In 2012, ReCAPTCHA started employing photos of house number taken from Google's Street View projects [1].

2. CAPTCHA
CAPTCHAs are automated tests designed to distinguish between human & computers by presenting with a problem that can't be solved by current computer programs [2]. Because CAPTCHAs can differentiate between bots & human with high probability, they are mainly used for many different security applications i.e. preventing bots from continuously voting in online polls, auto-registering for number of spam email accounts, automatically online shopping or to buy out an event ticket, etc. Once, CAPTCHA is cracked (i.e. computer programs/bots can successfully pass the test), they can impersonate humans & gain access to miscellaneous services that they should not. Therefore, it is important for CAPTCHAs to be secure [2].

2.1 Text based CAPTCHAs
It is the simplest form of CAPTCHA. The simplest yet innovative method is to present the user with some queries, which only a human being can answer.

Samples of such queries are:

1. What is forty plus four?
2. What is first the letter in ACADEMIA?
3. If today was a Monday, what is yesterday?

Such queries are simple for a human being to answer, but it's very problematic for bots to answer them. These are also pleasant to populace with visual disability – such as those with color blindness [2]. Further text CAPTCHAs contains text distortions and the user requested to recognize the hidden text.

2.2 IMAGE based CAPTCHA's
In this type CAPCHA the users has to identify image by performing image recognition task. First image CAPTCHA which used named as ESP Pix and it was developed at Carnegie Mellon University. A picture of ESP Pix CAPTCHA is shown in Figure 1. In ESP Pix CAPTCHA user has choose one image and to pass the test user has to select related word from a list of 72 choices

Fig 1: ESP Pix CAPTCHA

2.3. IMAGE based ReCAPTCHA

To preserve ancient books and other texts written before the computer age are currently being digitized to preserve human knowledge and to make information more accessible to the world. The pages are photographically scanned and transformed into text files by using optical character recognition (OCR) software.

This conversion into text is beneficial because the digitized text can then be indexed, searched, and stored in required format that can be easily analyzed and manipulated. One of the biggest problems in the digitization process is that OCR is inefficient at converting the words in images of scanned texts.

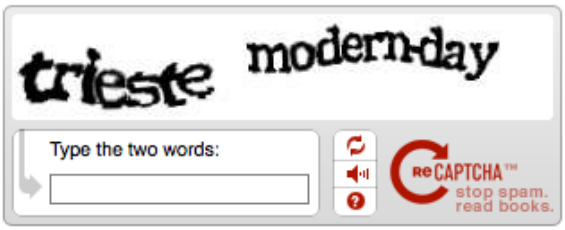

Fig 2: Image ReCAPTCHA

2.4 AUDIO based CAPTCHA

Image CAPTCHAs rely on superior human perception, leading to CAPTCHAs impossible to solve by people with vision impairments. Audio CAPTCHAs that rely instead on audio perception were introduced as to solve this problem but are much more difficult for web-users to solve.

Most of the CAPTCHAs on the web today show the pattern: the user is presented distorted digitized text and is asked to type the given text into a text box. The distortion technique is chosen so that it provides difficulty for automated bots to read the given text but can be easily possible for humans. In image CAPTCHA's, this most often means that graphic text is displayed with obfuscated characters. In audio CAPTCHAs, this means text is synthesized and mixed in with background noise, such as sound or unrecognized chatter [3]. Although these types of CAPTCHAs seem roughly analogous, the applicability of these CAPTCHAs is quite unalike due to inherent differences in the interfaces used to recognize and solve them.

Audio playback is linear. A user to solve an audio CAPTCHA first plays the CAPTCHA and then quickly concentrates to solve the text box [3].

2.5 AUDIO based ReCAPTCHA

Audio Based ReCAPTCHA has combination of essence of ReCAPTCHA and Audio Captcha. It enhances the functionality of ReCAPTCHA and Audio Captcha. The ReCAPTCHA nowadays are mostly used for digitalizing the texts which are not possible for Optical character recognizer to read, since the humans has more power of interpretation than that of computer OCR, similarly Audio based ReCAPTCHA will digitalize the audio [5]. The current drawbacks of speech to text conversion system are that it's impossible to distinguish between the accents of the voice, such as the word "Hello" spoken in French accent varies from Japanese accent. So mostly it happens that the voice to text converter can't recognizes the voice or understands it & the other problem is that it's not possible to convert the audio files which have noise with a high level of accuracy [4]. But on the other hand we humans can interpret the audio even with a high degree of noise. So it's fine to use human's fine hearing capability to do so. Below image (see Fig. 3) shows us a view of Audio based ReCAPTCHA.

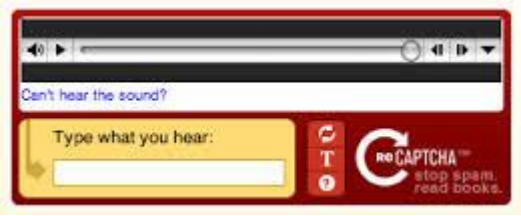

Fig 3: Audio ReCAPTCHA

When user fills the form he will get a CAPTCHA which is used for the security measure. Where user hears an audio file & type in what he hears. The typed data will be stored and mapped to the audio file. Based on the previous text mapped with the audio file the user will be shown a suggestion box [4]. The system feeds dictionary as well as we want to eliminate inefficiency caused mostly by spelling mistakes. Examples, if the user wants to type "Bye" & ends up typing "Dye" they both are different words for sure. At the database text processing will take place and if a particular text passes a certain amount of occurrences then it will be concluded. Pointers will be used to show the strength of the words. Since conversion of audio to text by current converters is inefficient. Therefore, human capability is used to convert the audio files to text [4]. Humans are able to interpret the voice even if spoken in different accents. Along with the provided security to distinguish between bots the users will be able to digitalizing the audio files.

3. APPLICATION OF CAPTCHA

Free Registering through net forms: Billions of websites on net offer free registering to services such as Email services, Social Networking sites, Online Gaming, Web blogs etc. Unfortunately, web robots attack many web sites. Web bots are typical programs, which automatically registered number of e-mail accounts on the net deteriorating web space [3].

Online polling: Real Inspiration of CAPTCHA originated from an online poll enquiring, "Which is the best graduate School in computer Science?" Carnegie Mellon University wrote a program that voted for CMU thousands of times. The next day, students at MIT's Students wrote their specificprogram and this became a contest between the voting bots. Can the outcome of online poll be trusted? Not as long as it ensures that only and only humans can vote on poll [5].

Web crawler: It is a computer script that browses the net in a systematic, automatedroutine or in a methodical manner [1]. It offers reasonable resolution to web pages that we want to be excluding from index by search engines.

d. Phishing Attack: Phishing is trying to get info such as usernames, bank details, credit card details and passwords, by camouflaging as a reliable entity. Plausible solution for phishing attacks can be preventing by use of CAPTCHA [2].

4. STRENGTH AND WEAKNESSES OF CAPTCHA

4.1 Text Based CAPTCHA

Text based CAPTCHAs are the highest universally used CAPTCHA these are used on Internet and web applications, but there are some weaknesses. The Number of group of characters and digits are undersized. As characters and digits have narrow geometry (limited font families) so it is desirable to recognize them through Optical character recognition or

OCR Technique [4]. After the noise is combined with the text based CAPTCHA they regularly create a trouble in recognizing CAPTCHA's. Though a number of digits and alphabets have special shapes, but after they are distorted, it is turn out to be hard to identify them. This difficulty is the majority in Text based CAPTCHA.

4.2 Image based CAPTCHA

The benefit of using Image CAPTCHA is that pattern recognition of in image is a difficult AI Problem and thus it is not easy to break this test using pattern recognition technique. The weakness of ESP Pix CAPTCHA shown in figure 2 has summarized below: ESP Pix is available only in English language so the end user must know the English vocabulary but there are only twenty seven percent internet users are English speaking [5].

a) Most of the time objects recognition becomes awkward due to the ambiguity present in picture object. In the place of Turing test it has turn into almost an IQ test.

b) It creates a difficulty to user having low vision [4].

The weakness of MMC scheme is same as described for ESP Pix CAPTCHA because it is available only in English. Text labels that are set in the image object are in simple fonts therefore it is able to be recognized by OCR technique easily. In this case the probability of entering into a site is 1/4 that is 25%, since the number of choices are 4.

The probability of guessing an image has been increase in CAPTCHA The Dog shown in figure 4 method, the probability is (1/9*9*9) that is 1/729 or 0.00137, since one image is selected out of 9 images and it repeats 3 times. The disadvantage of using this CAPTCHA is that it creates an extra overhead on the server since it requires 3 extra round trips to server [6].

5. CONCLUSION

In this paper, we have studied over the various types of CAPTCHA. A brief review has been carried out on the different Captchas and ReCAPTCHA. Analysis and experiments are needed to improve the security of our scheme. It is advisable to use the state-of-the-art speech recognition techniques for evaluating the security of acoustic CAPTCHAs, as it significantly yields higher success rates for CAPTCHA solving ability than previously suggested classification methods and therefore more expressive of the security risk as our data is at a risk of a CAPTCHA. Using the well-known ReCAPTCHA, we can see that the characteristics of the scheme have not altered notably within the last 5 years; the chances of comparatively high success rates for automated attacks are founded. In addition, the degree of signal distortion has been increased for this scheme at the cost of reducing human accuracy, renderingthe trade-off between usability and security problematic.

6. REFERENCE

[1] Haichang Gao, Honggang Liu, Dan Yao, Xiyang Liu,An audio CAPTCHA to distinguish humans from computers

[2] E. Athanasopoulos and S. Antonatos. Enhanced captchas: Using animation to tell humans and computers apart. In IFIP International Federation for Information Processing, 2006.

[3] L. von Ahn, M. Blum, J. Langford, "Telling Humans and Computers Apart Automatically Communications of the ACM", vol. 47, no. 2, pp. 57-60, Feb. 2004.

[4] J. of G. R. in C. Science and J. 2012,"Survey on CAPTCHA systems".

[5] Elie Bursztein, Steven Bethard, Celine Fabry, John C. Mitchell, Dan Jurafsky, How Good are Humans at Solving CAPTCHAs

[6] L. von Ahn, M. Blum, N. Hopper, and J. Langford. CAPTCHA: Using hard AI problems for security. In Eurocrypt, 2003.