Multi-Layered CAPTCHA- A New Approach to Tackle Web Robots

Dayanand
Research Scholar, Department of Computer Science and Information Technology, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, Uttar Pradesh, India

Megha Saloni
Student, Dept. of Information Technology , HMR Institute of Technology and Management, Hamidpur, New Delhi, India

Wilson Jeberson
Professor, Department of Computer Science and Information Technology, Sam Higginbottom University of Agriculture, Technology and Sciences, Allahabad, Uttar Pradesh, India

Abstract **- "CAPTCHA" stands for Complete Automated Public Turning Test[1], is a security mechanism designed to differentiate between online bots and human. It is a challenged-response test for computing to determine whether the user is human or not. It is used to defend against the malicious bots programs. As use of Internet has been become vital issue and many web application facing a threat of web bots(Robots). Web Bots or, Robots is an automated script which executes over an Internet and occupy web space and increases the network traffic [6]-[7]. The complication with currently used CAPTCHA i.e., text-based CAPTCHA or, graphic-based CAPTCHA is that, it is it is troublesome to read even for the human and Image based or, voice based CAPTCHA has been broken many times. Multi-layered CAPTCHA is type of hybrid CAPTCHA, in which two layers are there. First layer is face-recognition CAPTCHA and the second layer is either image-based CAPTCHA or, text-based CAPTCHA or, audio-based or, graphic-based CAPTCHA. This paper discusses the existing CAPTCHA and multi-layered CAPTCHA.**

Keywords:- CAPTCHA, Text-Based/Graphic-Based CAPTCHA, Image-Based/Audio-Based CAPTCHA, Face-Recognition CAPTCHA, Multi-Layered CAPTCHA

I. INTRODUCTION

CAPTCHA stands for Complete Automated Public Turning Test, is a security mechanism designed to differentiate between online bots and human. In a website, if a person wants to sign up for free e-mail services, then before submitting the web forms, he/she first has to pass a test. The test is very easy and simple for human but for the online bots or, computer it is impossible to solve that test. It is a challenged- response test for computing to determine whether the user is human or not. It is designed to prevent automated attacks by requiring users to perform tasks that are relatively easy for humans but challenging for web bots. CATCHAs provide an additional layer of security and are frequently paired with account login system to prevent brute force password attacks. The term CAPTCHA was devise or, coined in 2000 by Luis Von Ahn, Manuel Blum, Nicholas J. Hopper (all are of Carnegie Mellon University), and John Langford(then of IBM).[6][5]

CAPTCHAs automatically generate and evaluate the test; this test is difficult for the computer or, online bots but easy for the humans. If the success rate of human for solving the CAPTCHA is 90% or, higher than the computer programs or bots, only achieve a success rate of less than 1%[5]. So, the CAPTCHA can be considered a secure. There are some properties defined in development of CAPTCHA:[20]

- Automated:- Computer program should be able to generate the tests.
- Open:- The underlying database(s) and algorithm(s) used to generate and grade the tests should be public. This is accordance with Kerckhoff's Principle.
- Usable:- The effect of any user's language, physical location and perceptual abilities should be minimal.
- Secure:- The program generated test should be difficult for machine to solve by using any algorithm.

Currently, existing CAPTCHA implementation generally belongs to one or these categories, they are:-

- Text-Based CAPTCHA
- Graphic-Based CAPTCHA
- Image-Based CAPTCHA
- Audio-Based CAPTCHA

The most common CAPTCHA is Text-Based CAPTCHA, in which user have to enter the string of character that appear in a distorted form on the screen. Researcher had recently claimed that their simple generic attack have been broken a wide range of text-based CAPTCHA. The robustness of the text-based CAPTCHAs should rely on the difficulty of finding where each character is (segmentation), rather than what character is. So, the strong CAPTCHA have to be designed and built so that spammer cannot harm with web security. This paper migrates the shortcoming of existing approaches and proposed a new CAPTCHA, termed as Multi-Layered CAPTCHA, which is user friendly and add an additional layer of security to the existing CAPTCHA.

II. BACKGROUND DETAIL

The need for CAPTCHAs arises to keep out the website/search engine abuse by bots. In 1997, the AltaVista team comprised of Lillibridge, Adabdi, Bharat, began work on a system to prevent Internet bots from adding active URL's to the AltaVista, the search engine platform. To do this the AltaVista team work to prevent OCR (Optical Character Recognition), attacks by building puzzles and images which would cause OCR attack to fail. The AltaVista team worked to create system of varied typefaces, backgrounds, type style and size which would fool OCR reader [12]

In November 1999, slashdot.com released a poll to vote for the best CS College in the US. Students from the Carnegie Mellon University and the Massachusetts Institute of Technology created bots that repeatedly voted for their respective colleges. This incident created the urge to use

CAPTCHAs for such online polls to ensure that only human users are able to take part in the polls.
In 2000, Luis Von Ahn, John Langford coined the term CATCHA at Carnegie Mellon University (CMO). As, Yahoo's popular messenger chat service was affected by bots which pointed advertising link to annoying human user of chat rooms. So, the CAPTCHA was developed which is called as EZ-GIMPY which chose a dictionary word randomly and distorted it with a wide variety of image occlusions and asked the user to input the distorted word.[7]

ATTACKS OF CAPTCHA [5]
Challenge Reply Attacks:- If a CAPTCHA system can produce only a limited number of unique challenges, the automated agent may record all or, most of the possible challenges. The automated agent can then replay the correct answer whenever it is faced with the particular challenge for which it knows the correct solution. Some image CAPTCHAs are vulnerable to this weakness.
Bypass Attacks:- Any attack that overcome the need to solve the CAPTCHA at all. Generally, any system that sends the decode form of the CAPTCHA to the client program as a part of the data stream is vulnerable to such an attacks. Such attacks are not always a weakness of the CAPTCHA itself; they may instead be a weakness of the service using a CAPTCHA.
Single Processing Attacks:- The noise that are commonly used to obfuscate CAPTCHA images or, sound are intended to be one way; a computer should be able to add them but not reverse them easily. In principle, only a human flexible image and sound recognition capabilities should be capable of conveniently-reversing the transformations and recovering the original message.
Mechanical Turk Attacks:- Here, the problem of solving the CAPTCHA is automatically outsourced to a paid human agents. They immediately solve the challenges and quickly return the answer to the automated agent in the real time. The automated agent then presents the human-provided answer and is able to programmatically exploit the online resource.
Trivial Guessing Attacks:- If there is an unlimited range of challenges, but a very limited range of possible answer, then a high success rate may be achieved by an attacking program by merely guessing randomly from the available answer.
Brute Force Attacks:- If there is limited range of possible answer , then it is possible for the distributed group of automated agents to attack the CAPTCHA by exhaustive trying answer at random or, according to a selected sequence(example, a numerical 4-digit CAPTCHA would be have 10,000 possible answers).
Hybrid Attacks:- It is possible to combine these attack. For example:- If a signal processing attack which estimate 5 to 6 CAPTCHA character with high degree of confidence, a guess may be made on the remaining character yielding a success rate of between 1.5%(mixed case alphanumeric characters) and 10%(numeric digits).

Proposed CAPTCHA

Multi-layered CAPTCHA is the collaboration of Face-Recognition CAPTCHA and Text-based CAPTCHA or, Graphic-based CAPTCHA or, Image-based CAPTCHA or, Audio-based CAPTCHA. It comprises of two layers:

►At first layer, Face -Recognition CAPTCHA is implemented.

►Second layer would be either Text-based CAPTCHA or, Graphic-based CAPTCHA or, Image-based CAPTCHA or, Audio-based CAPTCHA. In this layer, the type of CAPTCHA will be changing randomly at every login attempt.

In Multi-layered CAPTCHA, the first layer is static CAPTCHA layer and second layer is dynamic CAPTCHA layer. Following types of CAPTCHA are present in this proposed CAPTCHA i.e., Multi Layered CAPTCHA:-
At first layer

Face-Recognition CAPTCHA:- This CAPTCHA presents users with a composite images containing several distorted human face along with some other objects and non-real face embedded in a complex background pattern. To prove that a user is human, user must solve the CAPTCHA by correctly selecting only the human face without choosing any other object or, non-real face image. If this is successfully done, the use is considered to be human and granted access to the secured resource.[9]

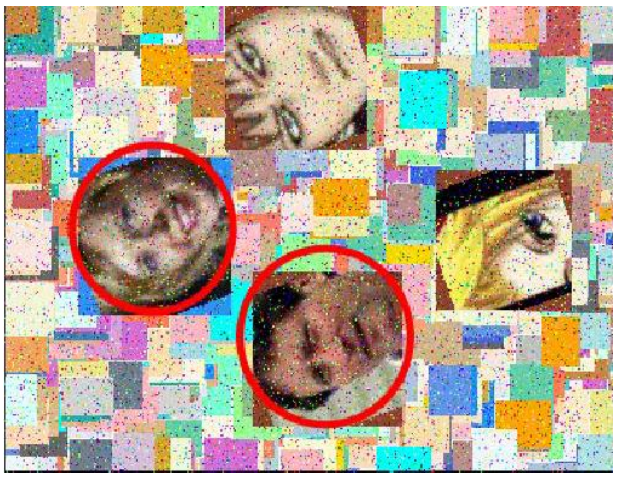

Fig 1: Face-Recognition CAPTCHA[9]

At second layer

- Text-Based/Graphic -Based CAPTCHA:-The most commonly used CAPTHAs is Text -Based CAPTHA where distorted text is displayed and used must recognize the distorted character and correctly enter them at designated space. Text-Based CAPTHA are of different types, they are as follow:-[7]

➢ Gimpy and Ez-Gimpy:- Gimpy is very reliable text CAPTHA built by CMU in collaboration with Yahoo for Messenger services. Gimpy works by choosing ten words randomly from the dictionary , and displaying them in a distorted and overlapped manner. Ez-Gimpy is a simplified version of gimpy CAPTHA, it randomly picks a single word from a dictionary and applies distortion to the text correctly.

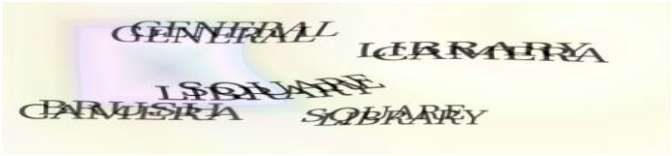

Fig 2. GimpyCAPTCHA[13]

Fig 3: Ez- Gimpy CAPTCHA

Baffle Text:- This technique overcome the drawback of Gimpy CAPTCHA. It was developed by Henry Baird at University of California at Berkeley. It pick up random alphabet to create non-sense but pronounceable.

Fig 4: Baffle Text CAPTCHA

- MSN CAPTCHA:- MSN CAPTCHA is used by the Microsoft under MSN umbrella. This is also called as MSN Passport CAPTCHA. In this eight character(upper case) and digit is used.

Fig 5: MSN CAPTCHA

- Bongo:- It is an example of graphic-based CAPTCHA, it is named after M.M. Bongard who published a book of pattern recognition problem in the 1979s. it displays two series of block. User must find the characteristic that sets the two series apart. User is asked to determine which series each four blocks belongs to.

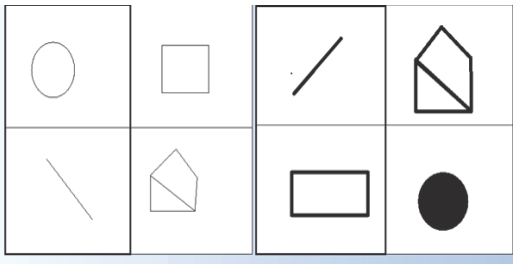

Fig 5: Bongo CAPTCHA

[1]. Disadvantage to Text-Based CAPTCHA:- The number of classes of characters and digits are very limited. When the distortion is added to text-based CAPTCHA, they often create a problem in recognizing them. In January 2008, article published an information week.com claiming Yahoo's [2]CAPTCHA., www.theregister.co.uk claiming that Google's[3] CAPTCHA has been broken by spammer.

- Image-Based CAPTCHA[7]:- In this CAPTCHA, the user is required to identify some images. The first image CAPTCHA is ESP Pix was developed by Carnegie Mellon University. In this user has given four images and in order to pass through the test, the user has to select word related to those four images from the drop-down list. Image-based CAPTCHA are of different types as, Asirra, CAPTCHA the Dog, Dynamic Image-Based CAPTCHA.

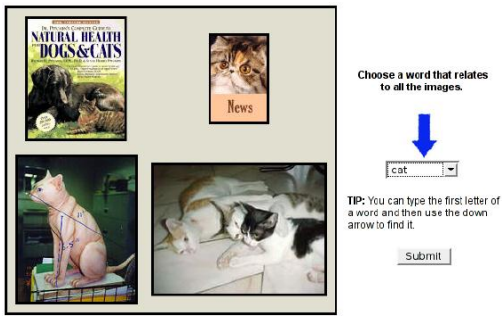

Fig 7: Image-Based CAPTCHA(ESP Pix)

Disadvantage of Image-Based CAPTCHA:- It creates a problem to user having low vision or, learning disability. Most of the time object recognition becomes complex due to ambiguity present in the image.

- Audio-Based CATCHA[7]:- In this type of CAPTCHA , a word or, a sequence of number is picked randomly. After that, render them into a clip using a TTS software and distort the audio clip and ask the user to identify and type the number or, word.

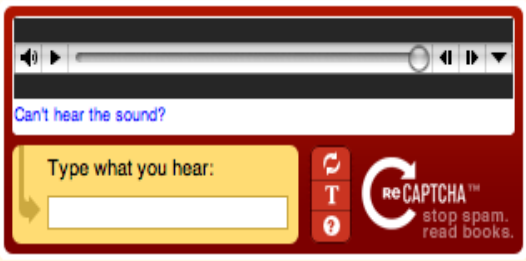

Fig 8: Audio-Based CAPTCHA

Disadvantage of Audio-Based CAPTCHA:- When noise or, distortion added to audio-based CAPTCHA, then it often creates a problem in recognizing the word. It also creates a problem to user having listening disability.

Advantages of Multi-Layered CAPTCHA

- As Multi-Layered CAPTCHA is consist of two-layer, so it is complex for the web bots to exploit this CAPTCHA.
- The first layer is Face-Detection CAPTCHA, on each login attempt, new CAPTCHA image is presented. So, the effective success rate of attackers is less than 1.6 in 1000[9].
- The second layer consist of Text-based CAPTCHA or, Graphic-based CAPTCHA or, Image-based CAPTCHA or, Audio-based CAPTCHA. On each login attempt, the' type of CAPTCHA' will be changes randomly.

- As, the second layer changes randomly so, for the web bots it would be difficult to solve this CAPTCHA and distort them but for human it would be easy.
- It is new comparison with existing CAPTCHA, so attackers are less vulnerable.
- Second layer adds an another layer of security to this CAPTCHA.

CONCLUSION

CAPTCHA is a challenged- response test for computing to determine whether the user is human or not. . As use of Internet has been become vital issue and many web application facing a threat of web bots(Robots). It is designed to prevent automated attacks by requiring users to perform tasks that are relatively easy for humans but challenging for web bots. CATCHAs provide an additional layer of security and are frequently paired with account login system to prevent brute force password attacks. CAPTHAs are of different types i.e., Text-based CAPTCHA or, Graphic-based CAPTCHA or, Image-based CAPTCHA or, Audio-based CAPTCHA. These CAPTCHA are more vulnerable to the attacks. In this paper Multi-Layered CAPTCHA has been proposed, which consists of two-layer. The first layer is Face-Detection CAPTCHA and second layer is Text-based CAPTCHA or, Graphic-based CAPTCHA or, Image-based CAPTCHA or, Audio-based CAPTCHA. On each every login attempt, the' type of CAPTCHA' will be changes randomly. As, the second layer changes randomly so, for the web bots it would be difficult to solve this CAPTCHA and distort them but for human it would be easy. This makes the Multi-Layered CAPTCHA more secure and less vulnerable from the attackers or, online bots.

APPLICATION of CAPTCHA

Following are the application of CAPTCHA:-

- In online Forms
- In Login Register
- In E-Ticketing
- Preventing Dictionary Attacks
- Prevent E-mail spam

REFERENCES

[1]. MoniNaor “Verification of a human in the loop or identification via the Turing test”.
[2]. Data electronically available at http://www.informationweek.com
[3]. Data electronically available at http://www.theregister.co.uk/2008/02/25/gmailcaptchacrack/
[4]. Data electronically available at http://blogs.zdnet.com/security
[5]. Haichang Gao, Mengyun Tang, Ping Zhang, Xiyang Liu "Research on Security of Microsoft's two-layer CAPTCHA"
[6]. Dayanand, Asha Pahuja Vishal Srivastav "SET BASED CAPTCHA- A NEW TECHNIQUE TO TACKLE WEB ROBOTS" in International Journal of advanced technology in Engineering and Science
[7]. Dayanand "WORD GROUPING CAPTCHA-A NOVEL APPROACH FOR SECURING WEB SERVICES" International Journal of Electrical, Electronics and Data Communication, ISSN (p): 2320-2084. Volume-1, Issue July-2013 pp. 10-14
[8]. ElieBursztein, Matthieu Martin, John C. Mitchell ”Textbased CAPTCHA Strengths and Weaknesses” in ACM Computer and Communication security 2011 (CSS’2011) Pages 11,12
[9]. Brian M. Powell, Gaurav Goswami, Mayank, Richa Singh, Afzel Noore "fgCAPTCHA: Genetically Optimized Face Image CAPTCHA 5"
[10]. Luis Von Ahn , Manuel Blum , and Jo n Langford “Telling humans and computer apart automatically”. In communications of the ACM Vol. 47, No. 2, February’2004, Pages. 57-58
[11]. Graeme Baxter Bell “Strengthening Captcha based web security” ,First Monday, Volume 17, Number 2 - 6 February 2012 Pages 2-3 http://firstmonday.org/ojs/index.php/fm/article/viewArticle/3630/3145
[12]. CAPTCHA COSC480 Spring 2014 (1).pptx
[13]. A Survey on CAPTCHA Categories Divyashree N , Dr. T. Satish Kumar