International Journal of Engineering & Technology, 7 (1.1) (2018) 256-259

SPC

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Web security through improvised image based CAPTCHA

R. Murugan, Tejasri. P *, NobeshReddy, G. Dinesh

Koneru Lakshmaiah Educational Foundation
**Corresponding author E-mail: tejapsri@gmail.com*

Abstract

CAPTCHA is a Completely Automated Public Turing Test to tell Computers and Humans Apart [1]. Whenever a website is hosted onto a server it is not human that always tries to access the website. Sometimes a human generated computer bot also known as Zombie may try to access the website. In such situations bots need to be filtered from legitimate users and this can be accomplished by using a simple method called CAPTCHA. Initially CAPTCHA was developed on text-based platform later it was evolved to audio, image etc. In this paper a new technique to differentiate bots from humans is introduced. It is a question-based image-oriented technique where a question comprising of keywords is posed along with four different images out of which only one image has relation with all the keywords. The user needs to spot the image to pass the test. A flag will be counting the number of times the user failed the test and depending on the number of failure attempts the user is judged.

Keywords: *CAPTCHA; Question based Image oriented technique; Image based CAPTCHA.*

1. Introduction

Exploiting World Wide Web is nothing but distribution of information and services which are confidential and valuable to the internet users. With the rapid growth of World Wide Web, the security becomes a major issue. To evade that issue, CAPTCHA is used. It is the acronym of “Completely Automated Public Turing Test to Tell Computers and Human Apart” which is universally a secure scheme to distinguished human from the bots. Numerous sites rely on CAPTCHA, to protect the assets from the bots or dangers. A Bot is the malicious programs a software application which has the capability to perform the repeated tasks automatically over the Internet. In recent years bot programs have been a major threat on the web [2]. The first inspiration for CAPTCHA originated from the online surveys and email spam.

CAPTCHA is a type of challenge response test and in other respects it is also called as reverse Turing Test. It was initially developed by Atla vista in 1997, to intercept bots from "including URL" function of their search engine. CAPTCHA is the program that ensures the web assets against bots by creating and evaluating the test that the human can easily solve though it's difficult to solve for the automated program.

There are many types of CAPTCHA like the text based, audio-based and image based. However, text- based CAPTCHA faces a problem of evolution of algorithms for analyzing the printed characters. Sometimes recognizing characters becomes difficult not only for the users but also for the bots. So, an image-based CAPTCHA is the solution for such difficulties. The most well-known CAPTCHA is an image of distorted text and numbers which have a great deal of disorder and noise in background. The user is requested to type the numbers and text which are in the distorted images show up on the screen. If the correct response is entered, then the system assumes that response is generated by a human or by the bots and the access is denied. The idea is that the legitimate user (human) can see through all the background fill and noise and it's difficult for bots.

1.1. Properties of CAPTCHA

Generally, the CAPTCHA should have the following properties.

1) It should be accessible.
2) It should be non-troublesome and straight forward to the end user.
3) It cannot stigmatize or redirect from or redirect from the basic role of the page.
4) It should be automated.
5) It should not put huge strain on the server or the browser.

2. Types of CAPTCHA

Some of the different types of CAPTCHA are mentioned below.

2.1. Text based CAPTCHA

Text-based CAPTCHA [2] is the most widely used CAPTCHA in web application. It is image of distorted text/numbers and addition to with some background noise or clutter. The content is generated randomly either text or alphanumeric. The user is asked to identify he distorted letters or numbers whatever displayed in the CAPTCHA challenge and entered them. It requires a large question bank. It is uncomplicated to solve for visual user, but it's becomes very difficult for blind user to read.

In text-based CAPTCHA simple asked question based on the arithmetic for example what is five plus three? Gimpy is a text-based CAPTCHA, randomly picked the words from the dictionary and are displayed in twisted, distorted manner. The users must type the three words from the challenge. Some of the examples of text-based CAPTCHA (Figure 1). E.g.-Gimpy is a simplified version of Gimpy. A word is picked from the dictionary and distorted, rotated and given to the user.

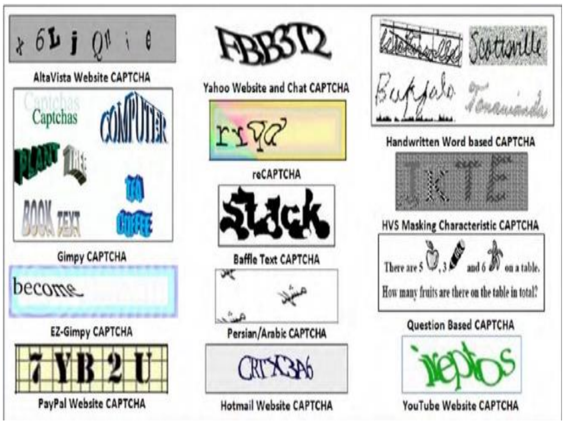

Fig. 1: Text Based CAPTCHA.

2.1.1. Text based CAPTCHA limitations

As this CAPTCHA implementation requires wide question bank. Through dictionary attack it is possible to crack this implementation. The concept involves, whenever a word is posed as question to the user. The user searches dictionary trying all the possible words to crack the CAPTCHA.

2.2. Audio based CAPTCHA

The Audio CAPTCHA [8] is based on sound. It is developed for visually impaired user. Audio CAPTCHA takes an arbitrary grouping drawn from recordings of words or numbers, consolidate them and include some disturbance/noise to it and given to the user. The user listens the sound clips and after that sort the talked word in the answer box and afterward submits it (Figure 2). Audio CAPTCHA is utilized to recognize the human from bots. Eco is the first audio-based CAPTCHA was actualized by the Nancy Chan from the City University.

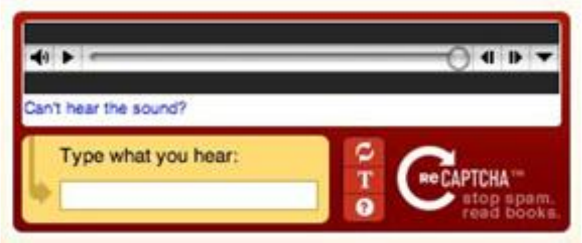

Fig. 2: Audio Based CAPTCHA.

2.2.1. Audio based CAPTCHA limitations

It uses classic radio program audio, which is not able to decipher by the automated speech recognition. It exploits the human capacity to understand words through context. The audio being utilized was initially recorded with the aim that it ought to be effectively seen by people. The audio CAPTCHA is slower than the visual CAPTCHA. It takes of an opportunity time to respond in due order regarding the test. To start with listen and replied in the answer box

2.3. Image based CAPTCHA

In the image based CAPTCHA[3], the user needs to pick those pictures that have the same resemblance.

ESP Pix is a first Image based CAPTCHA and it is accessible just in English language. It utilized a bigger database of images and animated pictures of regular items. The CAPTCHA framework gave a user an arrangement of pictures all connected with the same object or idea. The user was obliged to enter the object or idea to which all the pictures fit in with e.g. the system may present pictures of Globe, Volleyball, Planet and baseball anticipating that the user should accurately relate all these pictures with the statement ball. In general, image based CAPTCHA is nothing but a presentation of some simple logical pattern or some simple idea that can help to distinguish human and computer apart (Figure 3). The presented idea or pattern is constructed in such a way that it is easy for a human to solve or identify and difficult for a bot to do so.

It makes an issue to users having low vision or learning inability or because of smudging of pictures.

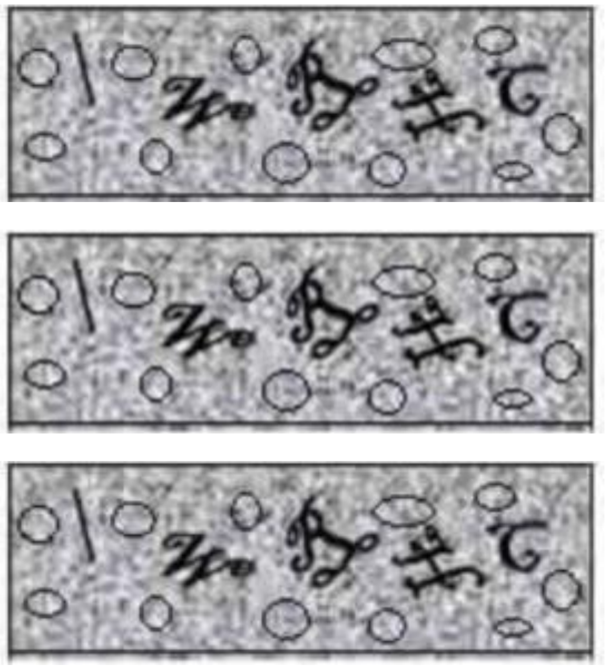

Fig. 3: Image Based CAPTCHA.

2.1.1. Image based CAPTCHA limitations

As imaged based CAPTCHA involves displaying actual images. An algorithm can be generated and trained to extract images from CAPTCHA each time and perform character recognition [9] with some simple image processing techniques (Figure 4). Later these images can be batched into different categories. The attack process involves, whenever an image-based CAPTCHA is displayed the algorithm is triggered and it checks the batched images for a perfect match. If there is a perfect match the CAPTCHA is cracked if not the algorithm stores the images and continues.

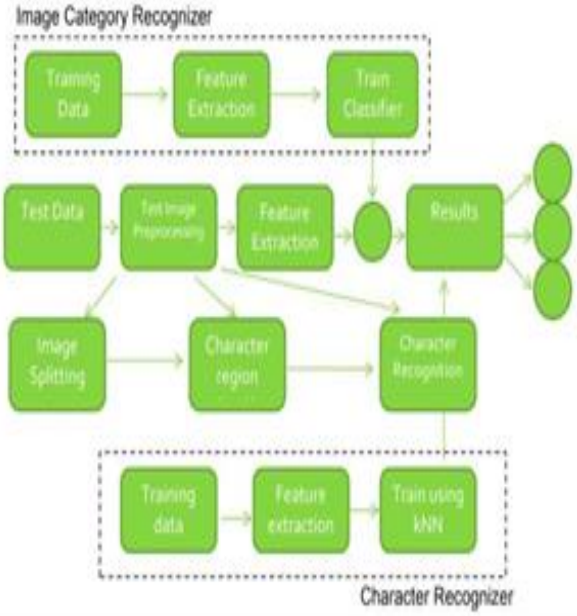

Fig. 4: Image Categorisation Algorithm.

3. CAPTCHA image generation process

STEP 1: Create a CAPTCHA image in two-dimension as given in Figure 3 because it is so much enough for holding characters that come out as an output of Linear Congruential Generator (LCG) [6].
STEP 2: Add some confusing techniques to uphold the characters and thus give some background color as dark granite and letters colors in black to confuse the user while finding out. Thus, these techniques help the human to find out easier because the colors are alike the RGB spectrum colors which are difficult for the computer programs and bots to find out.
STEP 3: we make the user confuse a lot as so to change the background and text color which will increase the rigidity of the user.
STEP 4: While designing CAPTCHA we need to give characters in small and non-confusable manner there are 16 different font style which will confuse the user and thus terminate the usage of bots.
STEP 5: Then the word set is divided into several pieces and each piece will have separate rotation with a random rotation value and the domain angle [-1, 1], [-3, 3], [-5, 5] in which it is rotated to confuse a lot the bots. The computer programs cannot find the word if it is dislocated at worst.
STEP 6: At last the worse more technique to confuse the computer program is to add random noise or disruption happened while entering the security tests. Thus, with all these enhancements in CAPTCHA it becomes more difficult for the bots to find out.

4. Proposed model

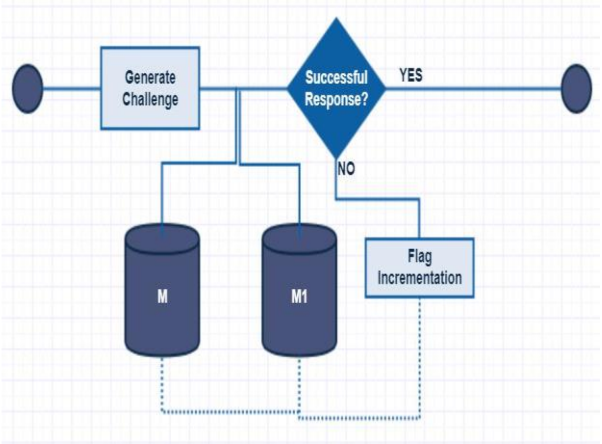

Fig. 5: CAPTCHA Generation and Working.

Our CAPTCHA implementation is a question base image-oriented model. It consists of two databases and a flag. The first database M stores all the text (words) which are to be posed as questions on the website when situation arises. The second database M1 stores all the images which are to be displayed on website in contrary with text. The flag keeps count of how many times the user failed the CAPTCHA test. Working of proposed model (Figure 5), whenever there is a situation of Distributed Denial of Service (DDOS) [10] attack or multiple requests from a single IP address the CAPTCHA challenge pops up. The challenge consists of a question comprising of keywords from database M e.g. “Yellow”, “Animal”,” Small” and below the question there will be a display of four images picked from database M1 in such a way that one among the four images have all the properties that are posed in the question. The selection of the keywords from the database M is done in such a way that for every set of keywords selected from database M there will always be an answer in the database M1. The selection of images from the database is done in randomly such that one among them is answer to question. Whenever the user fails the CAPTCHA challenge the flag is incremented and the test is refreshed with new set of keywords from database M and new set of images from database M1. If the value of the flag reaches certain threshold the user can be marked illegitimate.

4.1. Proposed model limitations

In this proposed model as, we are using two databases in order to store keywords and images which are to be displayed on the website as a CAPTCHA test. We require two highly efficient algorithms one for each database (M, M1). The algorithm operating on database M should be capable of picking keywords from database in a random way such that any intruder trying to crack the algorithm shouldn’t be able to find a logical link between the generated keywords. The algorithm operating on database M1 should be capable of picking images from database in such a way that there is always an image picked which has relation with all the picked keywords.

5. Conclusion

Text based CAPTCHA might be the most used CAPTCHA method, but it is not the most secure CAPTCHA methods. There are several attacks which cracked text-based CAPTCHA previously like Dictionary attack and pixel count. Audio based CAPTCHA are implemented very rarely because they have a lot disadvantages than advantages. Image based CAPTCHA are comparatively more secure than text-based and audio based but even the image based doesn’t live up to the mark. Attacks can be performed on image based also e.g. Brute force image storage. In the proposed model as, we are using a combination of text and images the probability of it being cracked reduces drastically. Theoretically it has the same complexity as image-based CAPTCHA, but it is more secure as the associativity of keywords with images is hard to crack. Also, flag is used to prevent repeated attempts from users up to a certain threshold.

6. Future scope

CAPTCHA is used to differentiate a human from a bot. It is used in many sites like health care, e-commerce sites, inventory management sites, business applications, supply chain Management and many more sites. It also reduces spam and virus attacks. So, rather than using a text-based CAPTCHA which has some limitations like distorted text and overlapping of letters it is better to use Image based CAPTCHA. Some of the advantages of Image based CAPTCHA are protecting website registrations, when compared to text- based image based increases the security. Image based decreases the complexity

References

[1] “The Official CAPTCHA Site.” [Online]. Available: http://www.CAPTCHA.net/.

[2] M. Tariq Banday, N. A. Shah, “A Study of CAPTCHAs for Securing Web Services”, IJSDIA International Journal of Secure Digital Information Age, Vol. 1. No. 2, December 2009.

[3] CaoLei,” Image CAPTCHA Technology Research Based on The Mechanism of Finger-Guessing Game”, School of Computer & Communication Engineering, University of Science & Technology Beijing, Beijing, 10083, P. R. China.

[4] CaoLei,” Image CAPTCHA Technology Research Based on The Mechanism of Finger-Guessing Game”, School of Computer & Communication Engineering, University of Science & Technology Beijing, Beijing, 10083, P. R. China.

[5] R.P. Anto Kumar, R. Sivakumar and S.S. Aalin Grace, “A New Implementation of Graphical Password Scheme for CAPTCHA Based Security System”, Middle-East Journal of Scientific Research 23 (7): 1353-1357, 2015.

[6] A. Clementeena and P. Sripriya, “Comparative Study of Algorithms used in CAPTCHA’s and New Finding Set as LCG Algorithm”, Indian Journal of Science and Technology, November 2016. https://doi.org/10.17485/ijst/2016/v9i42/97758.

[7] Ibrahim FurkanInce, IlkerYengin, YucelBatu Salman, “designing CAPTCHA algorithm: splitting and rotating the images against ocr’s”, Third 2008 International Conference on Convergence and Hybrid Information Technology, Busan, Republic of Korea.

[8] “A Study of Audio CAPTCHA and their Limitations” K. Aiswarya,. S. Kuppusamy, Department of Computer Science, Pondicherry University, Puducherry, India, Interna tional Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.

[9] “A Study of Audio CAPTCHA and their Limitations” K. Aiswarya, K. S. Kuppusamy, Department of Computer Science, Pondicherry University, Puducherry, India, Interna tional Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.

[10] COMS 6998 Computational Photography Spring 2009 “BREAKING AN IMAGE BASED CAPTCHA” FINAL REPORT Michele Merler, Jacquilene Jacob.

[11] “Detection of Distributed Denial of Service Attacks in Software Defined Networks” Lohit Barki, Amrit Shidling, Nisharani Meti, Narayan D G and Mohammed Moin Mulla, 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI), Sept. 21-24, 2016, Jaipur, India.