IJETCR
Open Access
ISSN: 2348 - 2117
International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Volume 2, Issue 4, Page No. 33-39

EVALUATING THE STRENGTH OF COGNITION BASED CAPTCHA BY THE APPLICATION OF CAPTCHA BREAKING ALGORITHM

Mrs. Geeta Gandhi [1], Dr. Vijay Dhaka [2]

Research Scholar, Jaipur National University, Jaipur, Rajasthan, India

gandhigeeta3@gmail.com[1]

Professor (Department of Computer Science, Jaipur National University, Jaipur, Rajasthan, India

Vijaypal.dhaka@gmail.com[2]

ABSTRACT

To secure the web services from automated spam bots nowadays CAPTCHA has become a standard security mechanism. It is a class of human interaction proof (HIP) which ensures that only human being is interacting with the system and not a bot at all. This ever increasing demand of secured web service motivates us to do more and more research in this area which leads to the generation of secured and robust CAPTCHA. Preceding in the same direction the present paper introduces a more reliable and innovative CAPTCHA based on cognitive ability of human. This CAPTCHA utilizes some of the designs created using alphabets and numbers to be used as CAPTCHA code. So, the major strength of this CAPTCHA code depends upon the design chosen. Further, to evaluate the strength of this CAPTCHA, the visual CAPTCHA breaking algorithm is applied to it to justify the robustness of present CAPTCHA generation technique.

I. INTRODUCTION:

The Completely Automated Public Turing Test to tell computers and Human Apart (CAPTCHA) are often the first line of defense in many online services [1].Almost all websites nowadays rely on HIP to limit the abuse of online email services , online shopping blogs and other online membership etc. .So, currently one of the most popular methods used to authenticate the existence of human operator over an unsecured connection is the CAPTCHA [2]. CAPTCHA is a means to limit the ability of attacks to scale their activities using automated bots [3]. The purpose of CAPTCHA is to distinguish between a computer and a human by presenting a challenge that is easy for most humans, but difficult for computers [4].

In most common implementation, a CAPTCHA consists of a visual challenge in the form of alphanumeric characters that are distorted in such a way that the available computer vision algorithms have difficulty in segmenting and recognizing the text. At the same time humans with some effort, have the ability to decipher the text and then respond to the challenge correctly [3] for e.g. EZ Gimpy currently used by Yahoo and Gimpy are CAPTCHA based on word recognition in the presence of clutter [5]. They are sometimes called as "Reverse Turing Tests" because they are intended to allow a computer to determine if a remote client is human or not [6].Thus, to avoid imposing undue user friction, a CAPTCHA must be easy for humans and difficult for machines [7], but due to the advances in machine learning the designing of such test is becoming increasingly difficult. So, while designing a defensive CAPTCHA technique, a CAPTCHA system designer should give consideration both to computer security and human friendliness.

Thus, there is a need to think beyond the existing CAPTCHA to provide a new and optimum solution to network security which can balance between security and user friendliness.

The objective of the present paper is here to introduce a new innovative CAPTCHA generation technique utilizing the cognitive ability of human being. It is based on some designs created using alphanumeric characters to be chosen as CAPTCHA code. Also, after introducing this new technique we measure the strength of this CAPTCHA by applying the visual CAPTCHA breaking algorithm to it and show that it is unable to break the innovative development.

II. BACKGROUND:

The term "CAPTCHA "was first introduced in 2000 by Van Ahen et al [3] which described a text that can

differentiate humans from computers to thwart automated attacks [9].

The Academic research into CAPTCHA takes the form of a friendly 'arms race' , with some researchers acting as 'malicious users' that try to attack and defeat the latest CAPTCHA systems automatically , while other researchers seek to design new defensive CAPTCHA techniques in response to known or anticipated attacks.[10]. In this way, we are grooming in both directions of web services as well as towards the development of artificial intelligence and machine learning. Further, we can categories the present available CAPTCHA as text based, image-based or audio-based CAPTCHA depending upon the type of test presented to the user for the access to web resources.

All these available types of CAPTCHAs are used to prevent corruptive usage of these services by automated scripts and thereby mitigate the threads of malicious attack [11].

Conclusively, we can say that CAPTCHA are based on the three principles as developed by Chew and Tygar [11];

i. Easy for humans to solve.
ii. Hard for computers to solve.
iii. Easy to generate and verify [11, 12].

Thus, if our website needs protection against abuse, it is recommended that our CAPTCHA must have the above qualities.

Also, while designing the defensive CAPTCHA technique a good CAPTCHA system should give consideration both to computer security and human friendliness which ensures an optimum CAPTCHA solution.

The next section introduces the new Innovative technology to generate CAPTCHA based on cognitive ability of human which further motivates to maintain a balance between usability and security proceeding in the direction of optimum CAPTCHA solution.

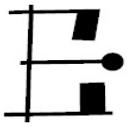
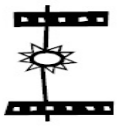
III. CAPTCHA: BASED ON COGNITION

By definition cognition means the mental action or process of acquiring knowledge and understanding through experience and senses. It is a set of abilities, skills or processes that are part of nearly every human action which can be utilized to make computers and humans apart which further leads to the objective of web security. Instead, cognitive science literature abounds with studies on visual perception showing that, for the most part, people do not require noticeably more processing time for object categorization comparative to an automated machine [13].

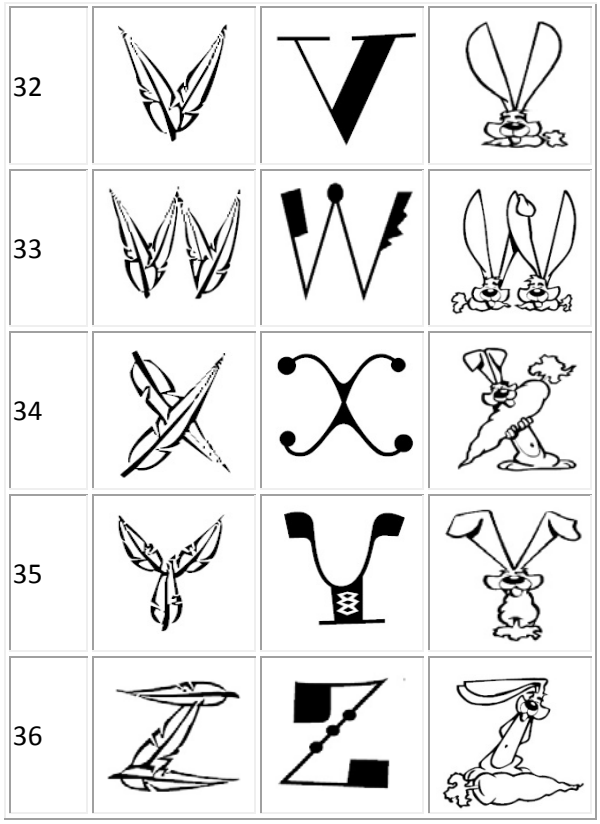
Based on this cognitive ability which can be utilized as a tool to differentiate a human from a machine we have generated a new Innovative CAPTCHA code. This CAPTCHA code is created by using alphabets and numbers to be hidden within any of the designs. In other words, we have taken a single number or alphabet created a design such that the specific alphanumeric letter becomes hidden between those designs, some of these developed design are shown in table (1) below.

Table 1: Designs based on cognition

S. No.	Image 1	Image 1	Image 1
1			
2			
3			
4			
5			
6			
7			
8			
9			

10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			

21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			



Now, after designing such patterns we can choose any of the existing CAPTCHA algorithms to display these designs on the screen.Further, the strength of this CAPTCHA depends upon the pattern or design chosen as a test, so for analyzing and evaluating the strength of present CAPTCHA code we have implemented the visual CAPTCHA breaking algorithm on the present developed code in the next section.

This algorithm developed by A.A.Chandavale, Dr. A.M.sapkal and Dr. R.M. Jalnekar who are the members of IEEE is for breaking the existing visual CAPTCHAs. On the application of this algorithm on our proposed CAPTCHA, we not only be able to evaluate and analyze its strength but also prove that our CAPTCHA is optimum source for web security.

The next section here is to apply the visual CAPTCHA algorithm to the CAPTCHA generated utilizing cognitive ability of human.

***IV. Application of visual CAPTCHA breaking algorithm on CAPTCHA developed using cognition*:** The more and more CAPTCHAs available nowadays, the more and more algorithms are developed to break them. The strength of a specific CAPTCHA can be estimated through the application of these algorithms on the CAPTCHA to be examined. Preceding in this direction here the VISUAL CAPTCHA breaking algorithm is applied on the CAPTCHA mentioned in this paper.

This algorithm consists of different phases mentioned in Fig (1):

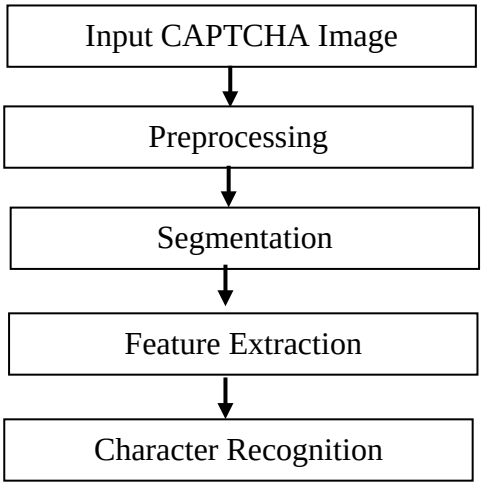


Figure 1: System Architecture

As mentioned in fig (1) the algorithm has different phases such as preprocessing, segmentation, feature extraction and character recognition. These steps or phases are now applied to the mentioned CAPTCHA for analyzing its strength. For this the CAPTCHA image will be given as an input to the program.

1. *Preprocessing* : Preprocessing will convert an input image into a cleared image by first converting it into grayscale, then carry out Binarisation, then removes lines , & dots(if any present on image).

The first step towards preprocessing is to obtain the given CAPTCHA image detached with the background. The removal of noise is the method which comes in the category of preprocessing. The preprocessing steps are explained below:-

A. *Single (No mesh) Background*

This step can be further performed in two steps

i The given CAPTCHA is converted into gray scale.

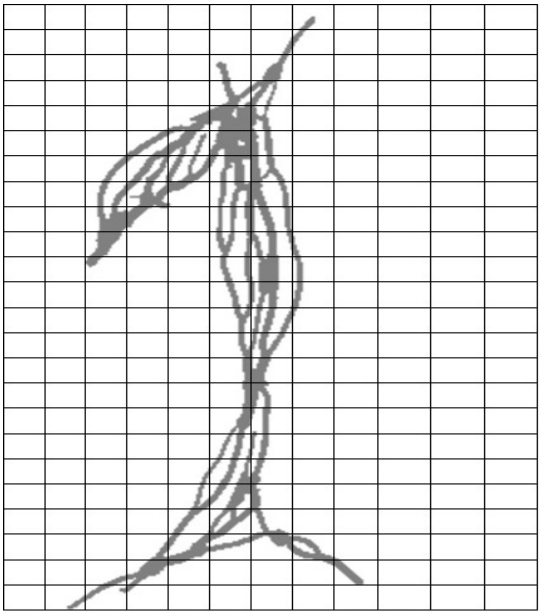
ii Binarisation of the image is performed as it can lead to an image whose pixels have only two possible intensity values.

(i) Color to gray scale: In this step the CAPTCHA image given to the program is converted into gray scale. In the case of mentioned CAPTCHA we have used only a single color to develop the CAPTCHA code so there would be no impact on the CAPTCHA code developed after converting into gray scale. In other words, our mentioned CAPTCHA would be unaffected through this step.

(ii) Binarisation: The binary images are the one whose pixels have only two possible intensity values namely 0 and 1. They are normally displayed as black and white

which are produced by Thresholding a color image in order to separate an object from background.

In our design as there is no background, only an image is there so Binarisation of the image would not be able to separate the background from the object. For e.g. consider the design of numeric 1 (one) as shown below in fig (2):



(a) CAPTCHA image with grid

0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	1	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	1	1	1	0	0	0	0	0
0	0	1	1	1	1	1	0	0	0	0	0
0	0	1	1	1	1	1	0	0	0	0	0
0	0	1	1	1	1	1	0	0	0	0	0
0	0	1	1	0	1	1	0	0	0	0	0
0	0	1	0	0	1	1	1	0	0	0	0
0	0	1	0	0	1	1	1	0	0	0	0
0	0	0	0	0	1	1	1	0	0	0	0
0	0	0	0	0	1	1	1	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	0	1	1	0	0	0	0	0
0	0	0	0	1	1	1	0	0	0	0	0
0	0	0	0	1	1	1	0	0	0	0	0
0	0	0	0	1	1	1	0	0	0	0	0
0	0	0	1	1	1	1	1	0	0	0	0
0	0	1	1	1	0	0	1	1	0	0	0
0	1	1	1	0	0	0	0	0	0	0	0

(b) Binarisation of Image

Fig (2) Binarisation of CAPTCHA code

In fig (2) it is clear that the Binarisation of the CAPTCHA code is not able to detach the numeric 1 from its image as the numeric 1 is itself a part of the image shown. This proves our CAPTCHA robust enough as even after Binarisation the machine is not able to detect the specific alphanumeric code hidden with the image, however it is very easy to be read by the human being.

Further Binarisation involves three steps as follows:

a) *Black mesh background*: - In black mesh background the removal of noise such as horizontal lines, vertical lines and dots that needs to removed to get clear image is performed, but as in the mentioned CAPTCHA there are no noise signals so this step has no use for us as it cannot affect our code in any manner.

b) *Line removal*: The CAPTCHA image sometimes contains horizontal lines and vertical lines. To remove these lines the number of continuous black pixel in row or columns is counted. If the count is more than 80% of the total width or height of the image, then it is detected as a line and removed by making it white. Now, as the mentioned CAPTCHA does not contain lines so there is no issue of line removal in this case.

c) *Discontinuity removal*: After the lines are removed, this step is performed to remove the discontinuity of the image by comparing it with the original image to fill the gaps in between to make the characters continuous. This step is again useless as there is no line removal in our CAPTCHA, so no discontinuity is generated which is required to be removed.

d) *Dot removal*: - After Binarisation sometimes the CAPTCHA may contain unnecessary set of black pixel i.e. Dots. To remove them, the image is scanned then, after getting the first black pixel we check its neighboring 8 pixels. If all of them are white, then make the black pixel white. This step is performed when there is a difference between the location of a white and a black pixel but as our CAPTCHA image is continuous this step would further lead to no change in the mentioned cognition based CAPTCHA pattern.

The next step towards preprocessing includes white mesh background

B. *White mesh background:*

This step further includes these steps:

i) The given CAPTCHA is first converted to gray scale and Binarized.

ii) The image is now inverted and the same mesh background as stated is applied.

iii) The image is now inverted to get back to the original image.

Any of these steps are not able to detach the alphanumeric code from our designed pattern so it is enough to predict the strength of mentioned CAPTCHA.

iv) Loosely connected characters: -I It is the last step towards preprocessing. It involves the noise removal after Binarisation. But, as there are no noise signals introduced in our CAPTCHA, we cannot differentiate between the mentioned CAPTCHA code with the image associated.

2. *Segmentation:* After the preprocessing stage the characters are required to be segmented so that they can become easier for the machine to detect and recognize. In our design the segmentation involves the breaking of the designs through different angles. In segmentation once the program checks black pixel, it is made red so that the program can understand that the specific character is already separated. In the mentioned CAPTCHA as there are patterns hiding the characters instead of the different characters closed to each other which requires segmentation to be recognized, so segmentation is useless in this case. Conclusively, segmentation is useless for the machine to recognize the specific letter which ultimately proves the mentioned CAPTCHA robust enough to secure the web services.

3. *Feature extraction*: The feature extraction is performed after the segmentation phase. It involves the recognition of characters based on the features or qualities of a specific letter or alphabet. The feature extraction phase makes it easier to recognize the characters through machine. The major features, which were used, are as follows:

a) *Number of Holes*: Each character is checked whether it has a hole or not. For e.g. Characters a, b, d, e, g, o, p, q each have a hole inside them so this feature of these letters can be utilized to recognize them with machine.

In the mentioned patterns, there may or may not have holes within the designs so it is unpredictable for the machine to recognize the specific alphanumeric character on the basis of holes. This proves the CAPTCHA strong enough to overcome the phase of feature extraction.

b) *Height of character:* - Each character is categorized into small or big on the basis of its height. For this a threshold is taken which characterizes the given character. In our CAPTCHA the height of the design involves the height of the object as a whole, so the height is not a appropriate method to identify the character involved within the certain design or pattern.

i) *Character recognition*:- The character recognition performed by comparing the readings like height of characters, number of holes, maximum number of white black transitions, nature of vertical stroke etc. calculated in the above steps. But, in our case as we are not able to evaluate any of the feature of the number or character which can be compared with the standard data to recognize the alphanumeric code presented as CAPTCHA, so it is hard to predict the exact code through the machine which further makes us successful towards the development of an optimum CAPTCHA.

V. CONCLUSION:

In this paper we have applied the visual CAPTCHA breaking algorithm to the mentioned CAPTCHA based in cognition. As observed that the above algorithm fails to break the innovative CAPTCHA which ultimately proves the strength of mentioned CAPTCHA technique. Further, it provides challenges to the field of Artificial Intelligence also leading to develop new doors in the fields of Technology.

VI. REFERENCES:

1. Manuel Egele, Leyla Bilge, Engin Kirda, Christopher kruegel, *"CAPTCHA Smuggling: thijaking web browsing session to create CAPTCHA"*, SAC' 10 March 22-26 2010, Sierre, Switzerland.
2. Simon R.Lang, Newille Williams, *"Impeding CAPTCHA Breakers with visual Decryption"*, Proc. 8TH Anstra Lasian Information Security Conference CAISC 2010, Bribane, Australia.
3. Marti Motyama, Kirill Lev Chenko, Chris Kanich, Damon McLoy, Geoffreg M. Vodker and Stefan Savage, *"Re CAPTCHA: Understanding CAPTCHA solving services in an Economic context"*, University of California.
4. Tianhui Cai, *"CAPTCHA solving with neural Networks"*, TJHSST computer system lab 2007-2008.
5. Greg Mori, Jitendra Malik, *"Recognizing Objects in Adversarial clutter: Breaking a Visual CAPTCHA"*, University of California, Berkeley.
6. E Lie Bursztein, Mathieu and John C. Mitchell, *"Text CAPTCHA Strengths and Weaknesses"*, Standford University.
7. E Lie Bursztein, Roman Beavxis, Hristo Paskov, Daniele Perito, Celine Fabry, John Mitchell, *"The Failure of Noise Based Non Continuous Audio CAPTCHAs"*, 2011 IEEE Symposium on Security and Privacy.
8. Shih-Yu Huang, Yeuan- kuen Lee, Graeme Bell and Zhan-he Ou, *"An Efficient Segmentation Algorithm for CAPTCHAs with line Cluttering and Character Warping"*, Department of Computer Science and Information Engineering.
9. Rich Gossweiler, Maryam Kamuar, and Shumeet Baluya *"What's up CAPTCHA? A CAPTCHA based on*

Image Orientation", www 2009, April 20-24, 2009, Madrid Spain, ACM 978-1-60558-4887-4/09/04.

10. Shih -Yu Huang, Yeuan- kuen Lee, Graeme Bell and Zhan-he Ou, *"A Projection-based Segmentation Algorithm for Breaking MSN and Yahoo CAPTCHAs"*, Department of Computer Science and Information Engineering, Ming Chuan University, R.O.C.

11. Christoph Fritsch, Michael Netter, Andreas Reisser and Gunther Pernul *"Attacking Image Recognition CAPTCHAs: A Naïve but Effective Approach"*, LNCS 6264, PP13-25, 2010.

12. *"CAPTCHA: Telling Humans and Computers Apart Automatically"*, www.captcha.net, The Official CAPTCHA Site.

13. Y. Xu, G. Reynaga, S. Chaisson, J. M. Frahm, P. Monrose and P. Van Oorschot *"Security and Usability Challenges of Moving-Object CAPTCHAs: Decoding Code words in Motion"*, Department of Computer Science of North Carolina at Chapel Hill, USA.