International Journal of Electronics Communication and Computer Engineering
Volume 4, Issue 3, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

Marque CAPTCHA- CAPTCHA Implementation Based on Motile Characters

Masarrat Mahedvi, Soumyashree Bilwar, Awani Joshi

Abstract **– CAPTCHA is a challenge-response test developed to ensure that the user is human and not an automated computer software program (known as bot). CAPTCHAs are used to prevent automated software programs or bots from performing actions that would degrade the quality of a given system.**

Internet today has occupied a major part in our everyday lives. Several services, including search engine, email and web board on internet are made available to the users without any cost indirectly making them extremely vulnerable. Bots are developed with purpose to use such services illegally and automatically. Therefore today almost all websites deploy CAPTCHAs on their sites in order to counter attacks from automated bots.

Unfortunately there are several bots available to break CAPTCHAs. Therefore several alternative methods of implementing CAPTCHAs have thus been developed but making it difficult for the users to decipher the code provided in CAPTCHAs.

We therefore propose a new system of CAPTCHA with an aim of providing enhanced security from the bots and at the same time being user-friendly. The proposed CAPTCHA system is based on marque type motile characters. As there is animation included in the form of motile characters, it will be difficult for the bots to attack this type of CAPTCHA meanwhile being easier for the users to decipher.

Keywords **– Bots, CAPTCHA, Characters, Frame, SHA.**

I. INTRODUCTION

CAPTCHA is a program that has been used for preventing the robot from accessing the information and resources. A framework was purposed to check the human users from bots in 1997, when search platform Alta-Vista sought a means of blocking automated URL submission to their engine [1]. The algorithm was developed by Alta Vista's chief scientist, Andrei Broder which was later on perfected by Luis von Ahn, Manuel Blum, Nicholas J. Hopper, and John Langford (all from Carnegie Mellon University) in 2000. They named it CAPTCHA which stands for "Completely Automated Public Turing test to tell Computers and Humans Apart". There are several web services and many other free services such as blogs, chat rooms, online polls, free e-mail accounts, search engines, password systems, etc., that are needed to be protected from malicious attacks by increasing the security. Since 2000, there are many types of CAPTCHAs developed to increase the level of security of CAPTCHA, however by raising the reliability its usability may decrease, so it a necessity to study existing reliability and usability elements of previous CAPTCHAs[2]. Although there are many problems to recognize some CAPTCHAs and many of them seem to be annoying and silly but in fact in terms of using CAPTCHA it is understood that CAPTCHA has good ability to protect websites and their resources from robotic attacks. Free email services are provided by many servers like Gmail, Yahoo, and Hotmail and so on. There is a need to protect user accounts from robots as they can hack accounts by implementing hacking algorithms. Also CAPTCHA has been used to prevent bots that create spam mail account in a huge numbers.

II. LITERATURE SURVEY

1. CAPTCHA based on spatial perspective and human imagination: In this type of CAPTCHA, user is required to rotate the given 3D model with the help of a slider till he gets the correct position i.e. position of slider at which the image is understandable to human. It is based on the imaginative ability of humans together with spatial perspective which helps humans to identify real object from random information background [3].

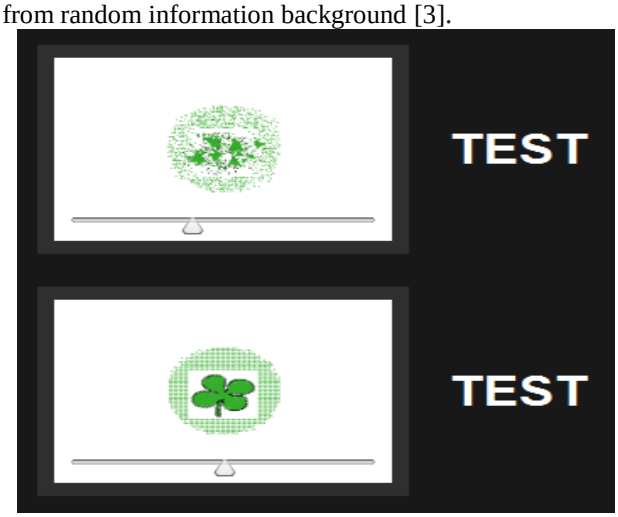

Fig.1. CAPTCHA based on spatial perspective

2. 3D CAPTCHA: This CAPTCHA consists of six alphanumeric characters which could be numbers or alphabets [4]. Each character in this type of CAPTCHA has its own axis and rotation angle. These characters are rotated around the axis with any random angle ranging from (-45^0 to $+45^0$)angle. This angle is selected by using any random function. In order to generate and adjust the sequence of characters in the CAPTCHA, ten factors were proposed:

• Rotation

• Overlapping

International Journal of Electronics Communication and Computer Engineering
Volume 4, Issue 3, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

• Obstacle such as straight line

• Distributed Noise

• Character Color

• Background Color

• Scaling

• Font

• Special Character

• Background Texture

Fig.2. 3D-CAPTCHA factors proposed

3. A Dynamic CAPTCHA Based on Persistence of Vision: This kind of CAPTCHA is based on the biological factor of humans of persistence of vision. The CAPTCHA consists of multiple frames which have useful as well as interfering characters. The frames containing the useful characters are static whereas the frames containing the interfering characters are dynamic i.e. constantly moving. When all the frames are switched quickly, humans can easily spot the useful characters that form the CAPTCHA due to persistence of vision. As the computers don't possess this quality of persistence of vision they won't be able to easily break the CAPTCHA [5].

Fig.3. CAPTCHA based on persistence of vision

III. TYPES OF CAPTCHA

CAPTCHAs can be categorized under four main categories:

1. Text-based CAPTCHA: The Text based CAPTCHAs are first of its kind which makes use of characters that are randomly generated from a specific database. It also makes use of some multimedia such as adding audio to the CAPTCHA that makes it difficult for the computers to recognize. To make sure that a human is accessing a website or a free email service, CAPTCHA is shown and the user would fill the empty box with the word in CAPTCHA. If user fill box correctly, user can continue using the service else he is denied access [6].

Drawback: Text based CAPTCHAs certainly do have many disadvantages as there are many algorithms written to break them.Several research projects have already broken real world CAPTCHAs, including one of Yahoo's early CAPTCHAs called "EZ-Gimpy" [7], CAPTCHAs used on e-banking sites, LiveJournal, phpBB [8] and on popular sites like PayPal [9].

Also PWNtcha or PWN CAPTCHAs project demonstrate the inefficiency of many CAPTCHA implementations and have successfully broken several CAPTCHAs used on popular sites with high efficiency [10]. The most popular method of breaking the CAPTCHA is based on optical character recognition (OCR) technology. Several papers based on OCR technique of breaking CAPTCHA have been proposed [7][13][14]. OCR is a technology which extracts the text from an image or a scanned document so that it can be edited, formatted, searched, automatically translated or converted to speech. The OCR technique is based on following method:

1. Pre-processing: This method includes techniques like line removal, layout analysis, zoning etc. to remove the distortion in the background in the CAPTCHA image.
2. Segmentation: Splitting the image into multiple regions in which each region contains a single character so that the image is much easier to identify.
3. Classification: Identifying the character in each region. There are different classification algorithms like Bayesian technique, linear classifiers like Fisher's linear discriminant and Naive Bayes classifier, Quadratic classifiers etc.

2. Image-based CAPTCHA: Image based CAPTCHA makes use of images to build a CAPTCHA thus making it difficult for the bots to recognize. It overcomes the limitation of text based CAPTCHA which is easier for the bots to break. But in image-based CAPTCHA because of having colors in all pixels and also having huge variety of meaningful images rather than texts and words, we will have better CAPTCHA [6].

Drawback: Even the image based CAPTCHAs are not full proof. Image based CAPTCHAs consist of users being shown a group of images based on which the user has to answer the query. By using image processing line detection or circle detection based methods single pictures are separated from the set of images given and characters hidden in the image are retrieved.Images are classified into different categories as required by the test. The characters corresponding to the images classified as being part of the required categories are then extracted. Another drawback of image based CAPTCHAs is that the sites implementing

International Journal of Electronics Communication and Computer Engineering
Volume 4, Issue 3, ISSN (Online): 2249–071X, ISSN (Print): 2278–4209

these CAPTCHAs would have to store a large database of images which may not be feasible for small sites. Color blind people may not be able to solve these CAPTHCAs based on color recognition.

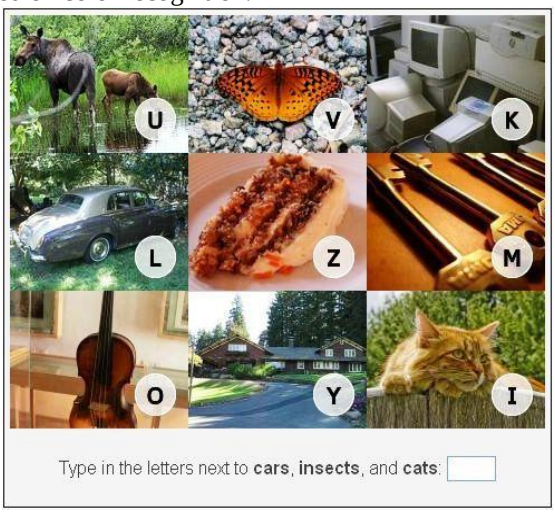

Fig.4. Vidoop-CAPTCHA

3. Audio-based CAPTCHA: Audio-based CAPTCHA was designed firstly because of some human disabilities such as disability of vision and other problems related to vision. The human who wishes to access the protected resource must identify the text that is displayed correctly. So users with vision related disabilities cannot solve CAPTCHAs and alternative called the Audio CAPTCHA has been introduced [6].

Drawback: In audio CAPTCHAs a computerized voice reads out letters or digits distorted by noise."Decaptcha" is able to break all the popular audio CAPTCHA schemes, including Microsoft and Yahoo [11]. Audio based CAPTCHAs also tend to annoy users by using high level of distortion. Besides this it cannot be bypassed by non-English users.Audio based CAPTCHAs have also been broken with high success rates [15][16]. The methods involved in the breaking include creating library of sounds representing each character in the character set of CAPTCHA. This might include creating several sounds for the same character as there might be distortions in the sound of each character. Then using voice-recognition software the audio CAPTCHA can be interpreted.

4. Video-based CAPTCHAs: This type of CAPTCHA makes use of a labeled YouTube video. In this type of CAPTCHAs a user must type some words that have best describe in the video [6].

NuCAPTCHA is a type of video CAPTCHA. It consists of video displaying characters that form the verification code. Video-based CAPTCHA is easy for the human to identify. The CAPTCHA consists of a short string of characters.

Drawbacks: ElieBurzten, a researcher at Google has proposed a scheme to break NuCAPTCHA. The algorithm consists of 5 phases:

1. Pre-possessing phase: In this phase the background images are removed and are converted into black and white to reduce comparison.
2. Frame Analysis: It is used to find objects that could potentially be CAPTCHA.
3. Cross frame analysis: This phase uses the result of the previous phase to isolate frames which contain actual CAPTCHA.
4. Segmentation: Segmentation aims at obtaining the characters.
5. Recognition: This phase is used to recognize characters.

IV. PROPOSED ALGORITHM

Step 1:

Initialize length of verification code. Select characters equal to the initialized length randomly from following character set :

{'A','B','C','D','E','F','G','H','I','J','K','L','M','N','P','Q','R','T','U','V','W','X','Y','1','3','4','6','7','8','9','@','$','*'}

The characters set like {'O', '0','2','Z','S', '5'}having similar appearance should been omitted from the character set to avoid confusion for the user. The selected characters form the verification code.

Step 2:

We set the screen size of the animation of *100 ×100* pixel. We consider a block of *5×5*pixel. As a result the animation screen has *20×20* blocks. Each block contains a character from the English alphabet set. Thus we see a frame having *20×20* alphabets as shown in Figure 2.These characters change randomly as the animation frame moves forward. Now we paint each character of the verification code on the animation screen separately.

Step 3:

Gray scale algorithm is applied on step 2 to convert the 24bit image of the verification code to an 8bit image. Gray scaling has been used to simplify the algorithm and reduce computational requirements.

Step 4:

Use thresholding to convert the image to a binary image.

Step 5:

Divide the frame into segments i.e. Blocks defined in step 2 and find out active pixels in every segment.Active pixels are the part of the characters which are to be displayed as the verification code. If a block contains an active pixel then the Alphabet in the block is highlighted.

Step 6:

Consider a 20*20 matrix. If a segment contains an active pixel consider '1' in the 0/1 matrix.

Step 7:

Construct a complete 0/1 matrix buffer and transmit to client and store the session ID and (Secure Hash Algorithm) SHA of CAPTCHA string i.e. the verification code on server.

Step 8:

In this step we draw the frame of animation. Each frame is 100 pixels* 100 pixels. We are using X, Y and Z plane for the movement of the frames.

The value of X and Y is a fixed, but the value of Z is determined by a hash function. The number of drawn frames is one greater than the number of characters in the

verification code. The additionally added frame is blank so as to indicate the end of CAPTCHA.

In the animation each point's value of coordinate Z is replaced by the adjacent right point's, and the points located in the rightmost column in the grid is replaced by the corresponding points' value of coordinate Z in the leftmost column in the same row.

Step 9:

Finally the generated CAPTCHA is shown in the Figure 5. Verification code characters are displayed in the frame. These frames containing the verification code keep scrolling horizontally like a marque.

Fig.5. Screenshot of proposed CAPTCHA

V. Results

Proposed CAPTCHA system is based on inclusion of 3rd Dimension to generate marque i.e. to scroll the verification codeand use of the characters to form the verification code. Special characters can also be included along with English alphabets and numbers in order to increase the scope. The proposed CAPTCHA frame is filled with letters assembled in row and columns. The verification code i.e. CAPTCHA characters are painted over this grid of letters. Screenshot of one of the frames of the CAPTCHA is shown in Figure 2.

The text-based CAPTCHA cracking algorithms are based on OCR (Optical Character Recognition) techniques. These CAPTCHA breaking algorithms are implemented on the snapshot of the CAPTCHA. Snapshot of the proposed system consists of alphabets in grid along with the verification code. Moreover the verification code rotates horizontally giving an animation effect. The charactersof the verification code are painted over the grid of alphabets which leads to confusion between the actual CAPTCHA and the background alphabets. Hence the proposed CAPTCHA cannot be easily cracked using a snapshot of the system by OCR CAPTCHA breaking algorithms.

In order to make it difficult for breaking, distortions in the image based CAPTCHA is very high. Due to which humans find it difficult to pass the CAPTCHA test leading to inconvenience. However in our proposed work, there is no distortion of the CAPTCHA image and the hence characters are clearly visible. Also, Image based CAPTHCAs require a large pool of images, thus requiring large amount of storage space. The newly proposed CAPTCHA has no images as it randomly generates the CAPTCHA code.

Audio based CAPTCHAs can be easily deciphered by voice recognition software. This CAPTCHA were basically introduced to help users understand the extremely distorted CAPTCHA text, as our system generates easily readable CAPTCHA an audio file is not necessary.

Proposed CAPTCHA consists of moving characters which looks like a video. Video CAPTCHAs are broken by segmenting the video snapshot similar to OCR methods. Due to the presence of grid of small characters segmentation fails to break the CAPTCHA.

Hence the proposed algorithm overcomes the shortcomings of the text-based, images-based, audio-based as well as video-based CAPTCHAs.

VI. Future Work

In the proposed work the letters which form the grid are all of same color. The CAPTCHA i.e. the verification code characters are also of the same color. Hence the CAPTCHA has only 2 colors as seen in Figure 2. There are many available segmentation and color-clustering algorithms which may pose a threat to our CAPTCHA design. In order to make the algorithm stronger each character of the CAPTCHA can be made of a different color. The color recognition algorithm removes all the extra objects i.e. the noise from the CAPTCHA and then the characters are recognized by segmentation. If the verification code is of different colors the breaking algorithm may get confused between the noise in the CAPTCHA and the actual verification code. In such a way thecolor recognition algorithms will also fail to detect the characters of the verification code. Thus as a future work the color of the different characters in the verification code can be made different or the colors of the entire verification code should change randomly at equal intervals.

VII. Conclusion

Proposed CAPTCHA can be used as a solution to existing problems. In early versions a text string was generated on a distorted background, which was difficult for the users to interpret. The new CAPTCHA will provide an animated version of CAPTCHA characters where the CAPTCHA code itself is made up of motile characters which will provide enhanced protection from malicious attacks of bots and at the same time be user-friendly.

VIII. ACKNOWLEDGEMENT

We wish to express our deep sense of gratitude to our guide, Prof. B. D. Phulpagar for his valuable guidance and useful suggestions.

REFERENCES

[1] N. Greene, (2012, July 11). "The Story Behind CAPTCHA". Available:http://www.pcmech.com/article
[2] R. Pakdel, N.Ithnin and M. Hashemi, 2011. CAPTCHA: A Survey of Usability Features. Research Journal of Information Technology, 3: 215-228.
[3] Juraj R., Zilina, (2010, Oct). "Captcha based on spatial perspective and human imagination". Available: http://www.3dcaptcha.net
[4] Imsamai, M.; Phimoltares, S., "3D CAPTCHA: A Next Generation of the CAPTCHA," Information Science and Applications (ICISA), 2010 International Conference on , vol., no., pp.1,8, 21-23 April 2010
[5] Tao Men, Xiuqiong Zhang, Guo Huang, Yan Sun, "A Dynamic CAPTCHA Based on Persistence of Vision," Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on, (23-25 March 2012) vol.1, no., pp.676, 679.
[6] hu, B.B., J. Yan, Q. Li, C. Yang and J. Liu et al., 2010. Attacksand design of image recognition CAPTCHAs". Proceedings of the 17th ACM Conference on Computer and Communications Security, October 4-8, 2010, Chicago, IL.USA. pp: 187-200
[7] Mori, G.; Malik, J., "Recognizing objects in adversarial clutter: breaking a visual CAPTCHA," Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, vol.1, no., pp.I-134,I-141 vol.1, 18-20 June 2003, doi: 10.1109/CVPR.2003.1211347.
[8] Shujun Li, Syed Amier Haider Shah, Muhammad Asad Usman Khan, Syed Ali Khayam, Ahmad-Reza Sadeghi and Roland Schmitz, "Breaking e-Banking CAPTCHAs," in Proceedings of 26th Annual Computer Security Applications Conference (ACSAC 2010, Austin, TX, USA, December 6-10, 2010), pp. 171-180, 2010, DOI: 10.1145/1920261.1920288
[9] Kluever, Kurt. "Breaking the PayPal CAPTCHA". Available:http://www.kloover.com
[10] S. Hocevar, Kaka Labs, "PWNtcha". Available: http://caca.zoy.org/wiki/PWNtcha
[11] Elie Bursztein, Romain Beauxis, Hristo Paskov, Daniele Perito, Celine Fabry, John Mitchell, The Failure of Noise-Based Non-continuous Audio Captchas, Proceedings of the 2011 IEEE Symposium on Security and Privacy, p.19-31, May 22-25, 2011.
[12] Wenjun Zhang; Zhang's CAPTCHA architecture based on intelligent interaction via RIA, 2nd International Conference on Computer Engineering and Technology (ICCET) 2010, vol 6.pp: 57 - 62.
[13] Chandavale, A.A.; Sapkal, A.M.; Jalnekar, R. M., "Algorithm to Break Visual CAPTCHA," Emerging Trends in Engineering and Technology (ICETET), 2009 2nd International Conference on , vol., no., pp.258,262, 16-18
[14] Yan, J.; El Ahmad, A.S., "Breaking Visual CAPTCHAs with Naive Pattern Recognition Algorithms," Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual , vol., no., pp.279,291.
[15] Tam, J., Simsa, J., Hyde, S., Von Ahn, L. "Breaking Audio CAPTCHAs". Available: http://www.captcha.net
[16] Elie Bursztein , Steven Bethard, Decaptcha: breaking 75% of eBay audio CAPTCHAs, Proceedings of the 3rdUSENIX conference on Offensive technologies, p.8-8, August 10, 2009, Montreal, Canada.

AUTHOR'S PROFILE

Masarrat Mahedvi
A final year Computer Engineering student currently pursuing her B.E degree in P.E.S Modern College of Engineering, Pune, India.
Email: masarrat.mahedvi@gmail.com

Soumyashree Bilwar
A final year Computer Engineering student currently pursuing her B.E degree in P.E.S Modern College of Engineering, Pune, India.
Email: bilwar.soumya@gmail.com

Awani Joshi
A final year Computer Engineering student currently pursuing her B.E degree in P.E.S Modern College of Engineering, Pune, India.
Email: awanijoshi@gmail.com