International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 08, August 2016
www.ijates.com
ijates
ISSN 2348 - 7550

DESIGN AND IMPLEMENTATION OF A USER FRIENDLY TEXT BASED CAPTCHA

Sapna Vij[1], Dr. Harish Rohil[2], Dr. Manju[3]

[1]*Assistant Professor Department of Comp. Sc. & Appls. Ch. Devi Lal University (India)*

[2]*M.Tech. Scholar Department of Comp. Sc. & Appls. Ch. Devi Lal University (India)*

[3]*Lecturer (Computer Engg.) Govt. Polytechnic, Nathusari Chopta, Sirsa (India)*

ABSTRACT

CAPTCHA is a Human Interactive Proof (HIP) system which is used to distinguish between human users and computer programs automatically. CAPTCHA has been widely used for preventing malicious programs to access web resources automatically. Web accessibility is increasing day by day with the advent of public services being online making space for payment gateways that further make space for data theft, information theft which is very curious. So, In this paper, a user friendly text based CAPTCHA is proposed for secure authentication. To evaluate the effectiveness of the proposed CAPTCHA, an attack to decode the CAPTCHA was made and compared with other existing real time CAPTCHAs being used at various websites.

Keywords: Security, Authentication, CAPTCHA, Reverse Turing Test, Text Based Captchas

I. INTRODUCTION

CAPTCHA stands for Completely Automated Public Turing Tests to Tell Computers and Humans Apart [11]. In 1997, the AltaVista team comprised of Lilli Bridge, Adabdi, Bharat and Broder, began work on a system to prevent Internet bots from adding active URL's to the AltaVista search engine platform. AltaVista was the first to use a simple CAPTCHA that generated images of random text [15]. The second team comprised of Luis von Ahn, Manuel Blum, Nicholas J. Hopper and John Langford gave birth to the term CAPTCHA in 2000 [5].They introduced it while working at Carnegie Mellon University.

The primary benefit of using the CAPTCHA system is that it lets distinguish between human being and computer program. It is very important and required because the computer & Internet services are meant for human being uses for their thoughtful purpose not for spreading rumors and unethical behaviors. CAPTCHAs are sometimes called “Reverse Turing Tests”, because they are intended to allow a computer to determine if a remote client is human or not [3].

CAPTCHA can be classified into 3 main categories:

- Text Based CAPTCHA,
- Audio/Sound Based CAPTCHA,
- Image based CAPTCHA.

- **Text Based CAPTCHA:** The text based CAPTCHA typically rely on sophisticated distortion of text images rendering them unrecognizable to the state of the art of pattern recognition programs but recognizable to human eyes [1].
- **Audio/Sound Based CAPTCHA:** Audio-Based CAPTCHAs are based on the sound-based systems. Audio-based CAPTCHAs ask users to recognize the vocabulary that is heard from a speech. It contains downloadable audio-clips. In this type of CAPTCHA, first the user listens and after that submits the spoken word.
- **Image Based CAPTCHA:** In image-based CAPTCHAs, users have to identify the subject of an image. This type of CAPTCHA usually interacts with users by using a pointing device, e.g., mouse [4]. Graphics-based CAPTCHAs are challenge-tests in which the users have to guess those images that have some similarity. For example: visual puzzles.

In this paper, the text based CAPTCHA is proposed because of its ease of use. Text-based CAPTCHAs are very easy to implement. It asks users to recognize the word, which has been presented in a distortion form. This type of CAPTCHA is intuitive to users [4]. In other words, it is easy to use without learning or training.

A good CAPTCHA must not only be user friendly but also robust enough to resist computer programs that attackers write to automatically pass CAPTCHA tests. It should be easy for humans to solve but next to impossible for machines to solve [9].A good CAPTCHA should satisfy two main aspects: robustness and usability.

The robustness is its strength to defend against adversarial attacks; whereas the usability is the ease with which humans pass its challenges [2].

The test should be automatically generated and graded by a machine. This is the main requirement of a CAPTCHA.

II. RELATED WORK

A lot of research has been made on CAPTCHAs by researchers. Based on the classification of CAPTCHAs, literature review can be summarized as:

2.1 Text Based CAPTCHAs

The literature related to text based CAPTCHA is given below:

Yan, J. and Ahmad [13] and Suliman A. Alsuhibany [2] discussed about the usability issues of CAPTCHA. Robustness and usability are two fundamental aspects with CAPTCHA. They discussed usability issues that should be considered and addressed in the design of CAPTCHAs.

A. A. Chandavale and Dr. A.M.Sapkal proposed the framework for CAPTCHA strength measurement under which an algorithm for CAPTCHA solver is implemented and implemented CAPTCHA solver (breaking) using modules like Preprocessing, Segmentation and Character recognition for Ez Gimpy CAPTCHA [16], [1].

Rituraj Soni and Divendra Tiwari improved the collage CAPTCHA method by presenting a method for increasing the resistance of it. Although represent that we can increase the rate of its difficulty in order to improve its resistance against the attacks by applying other effects such as increasing the objects present in the

screen and decreasing the distance between the objects, this way the test will become more difficult even for a human user in addition to computer programs [17].

Xiao ling-zi and Zhang yi-chun discussed about the design of CAPTCHA. The CAPTCHA implementation is tricky and risky without deliberate design and discussed the methodology of breaking the CAPTCHA and comparing the security of the various CAPTCHA designs by applying a certain approach and comparing the number of segments with total number of characters in the CAPTCHA [6].

2.2 CAPTCHAs based on HIP or Visual CAPTCHAs

Sushama Kulkarni [8] and A. Caine [12] explored that AI can be used for the designing of efficient CAPTCHA and we can increase the security of web based applications.

Kumar Chellapila and Patrice Y. Simard looked for the tasks where machine learning algorithms are not good as humans with the hope of gaining insight into their current limitations and found that most HIPs are pure recognition tasks that may easily be broken using machine learning. From this observation, he found that building segmentation tasks is the better way to confuse machine learning algorithms [18].

Henry S. Baird et. al. stated that the Scatter Type CAPTCHA, was designed to resist character–segmentation attacks and shown to be highly legible to human readers, is analyzed for vulnerabilities and is offered for experiments [19].

2.3 CAPTCHAs for Web Security

M. Tariq Banday and N.A. Shah reviewed the existing CAPTCHA schemes that have been proposed or are being used to protect various Web services. In this paper author discussed CAPTCHAs various security and usability issues in CAPTCHA design and provide guidelines for improving their robustness and usability by applying suggested design considerations such randomization of characters, auto sizing of the characters and distortions of the characters and adding random noise at foregrounds and at backgrounds that makes a CAPTCHA usable and secure [10].

Kanika Singhal generated robust and human friendly CAPTCHA. The CAPTCHA is generated using Markova text and time variance. Noise addition along with misalignment of characters is also done to increase the robustness of CAPTCHA [14].

III. PROPOSED WORK

Popular web sites are subject to brute force attacks by programs or computers known as Robots or Bots. They can be used to break user accounts or submit an unlimited number of service requests such as email account creation, web connection requests and serving as shopping agents. Such activities often lead to abuse of privilege causing the server to exhaust its resources or worse cause it to shut down. A key part in security is authentication means user or computer has to prove its identity to the server or client. So we proposed a CAPTCHA i.e. User Friendly CAPTCHA in which CAPTCHA is designed as a standalone CAPTCHA rather than in the form of a script which is written in scripting language i.e. PHP, Java script, VB Script etc. Proposed CAPTCHA generates such challenges that are hard to segment. After the designing of CAPTCHA attacks are made on sample set CAPTCHA image and designed set CAPTCHA image keeping length of images of both sets

same. Comparative analysis of CAPTCHA images is done by considering number of segments, turnaround time and problem solving accuracy with equal length of images of both sets.

3.1 Proposed CAPTCHA Design Process

The following Figure 1 describes about the process of proposed CAPTCHA design:

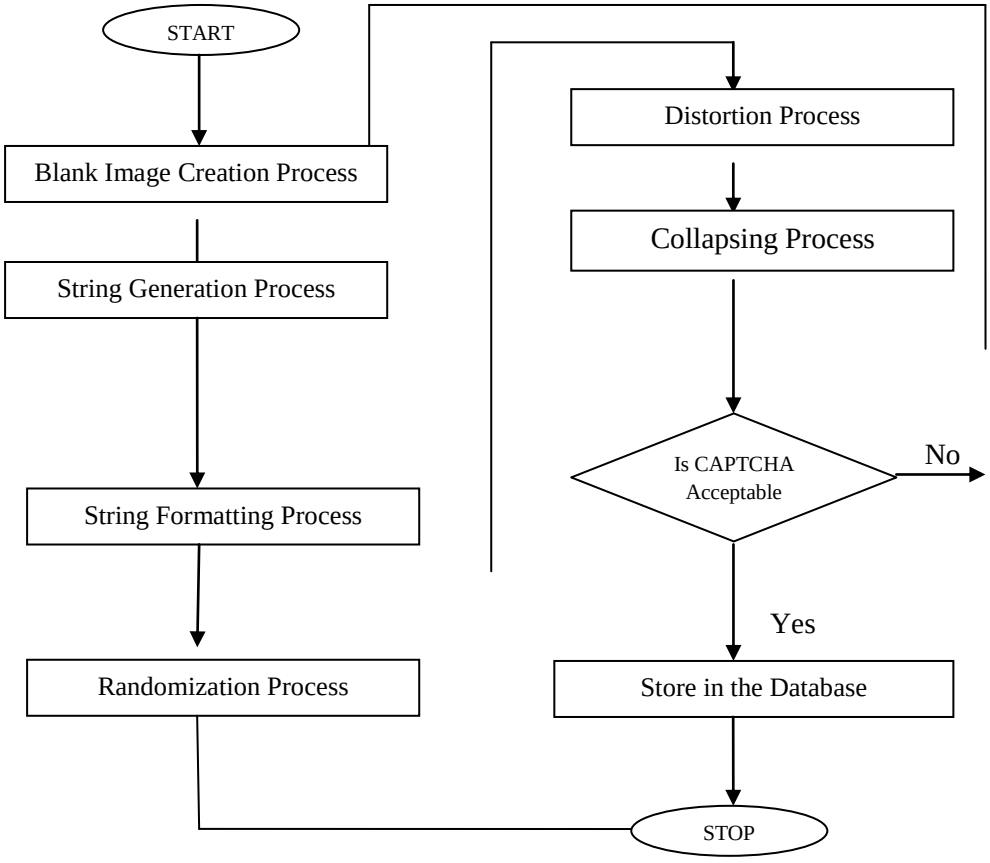

Figure 1 Proposed CAPTCHA Designing Process

3.1.1 Image Creation Process

This process involves the creation of bitmap image that will hold the string into it. This process maintains the size of the Bitmap according to the size of the string and adjusts the string into it.

3.1.2 String Generation Process

This process involves creating set of characters from a set of characters that will be going to emboss on the image and finally become CAPTCHA. In this process the size of string to be embossed on image is decided.

3.1.3 String Formatting Process

This process involves formatting the different sets of styles of the text and giving various formatting to the text string before finally embossing to the image.

3.1.4 Randomization Process

In this process all the design parameters are given random values and embossed on the image that will make the CAPTCHA image and it comprises of the anti segmentation techniques and anti recognition techniques that makes it harder to segment and recognize.

3.1.5 Distortion Process

In this process the actual transformations on the image and distortions is done in which string is auto resized as it grows and touching the boundaries and giving the certain angle in which it rotates.

3.1.6 Collapsing Process

In this process all the characters in the string are given a random spacing and look sometimes that it overlapping of the characters.

3.2 Assumptions for Proposed CAPTCHA Designing

The assumptions considered for design of proposed CAPTCHA are given below:

- The design of the CAPTCHA is done in VB.net and it may works in Visual Studio Environment only.
- Only bold upper English alphabets are used.
- It is assumed that if numbers of segments are less than numbers of characters in the CAPTCHA image then image cannot be recognized.
- The breaking of designed CAPTCHA and existing CAPTCHAs are tested in MATLAB 2010b version, we have tried to make sure that this proposed approach will work for all sizes and it will work only on text-based CAPTCHA in which there is certain gap must be there among each character.
- The proposed approach processes only jpeg images.

3.3 Algorithm for Proposed CAPTCHA Design

The algorithm for the designing of proposed CAPTCHA is given below:

Step 1: Create Bitmap at runtime to store the image.

Step 2: Random string with random no of characters is generated and warped on the image.

Step 3: The String text size of the image grows randomly and adjust automatically.

Step 4: The background of the bitmap is distorted using the anti-segmentation & anti-recognition techniques.

Step 5: Pass the string on the image to emboss it on the image to get final CAPTCHA.

Step 6: Check the designed CAPTCHA for its suitability.

Step 7: If it qualifies then store it in the database otherwise, neglect it.

Figure 2 shows some of the designed CAPTCHA images

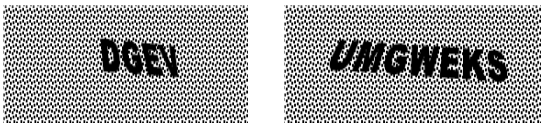

Figure 2 Designed CAPTCHA Images

To check the suitability and efficiency of proposed CAPTCHA, an algorithm was designed to break the CAPTCHA. This algorithm is also helpful in making comparison of the proposed CAPTCHA with other existing process of breaking CAPTCHA and CAPTCHAs. The process of breaking CAPTCHA and algorithm are given as under:

3.4 CAPTCHA Breaking Process

The overall breaking process of CAPTCHA goes through series of further process as follows and as shown in Figure 3:

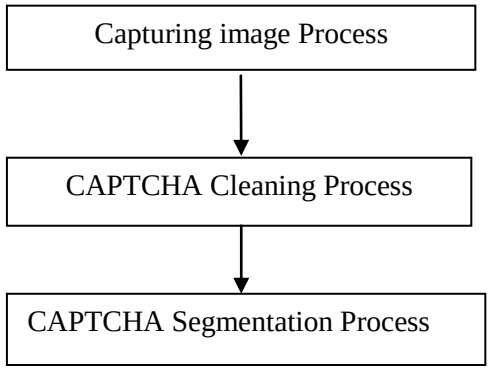

Figure 3 CAPTCHA Breaking Process

3.2.1 Capturing Image Process

In this process the CAPTCHA image is taken either from a database or designs your own using CAPTCHA designing guidelines.

3.2.2 CAPTCHA Cleaning Process

In this process the CAPTCHA acquired from the source is first converted to gray mode and then its compliment is taken to threshold the values such that value above 125 will be 1 and below values will be 0.

3.2.3 CAPTCHA Segmentation Process

Segmentation is the process by using a variety of means, parses the text pixels into separate character segments [13]. In this process the compliment of the cleaned image is taken and the objects are removed whose pixel values are less than 50. Then again take the compliment the image for better view of the image and then measure the properties of the image regions and then crop from the corners of the image and label them so that it could be noted that how many number of segments are there in the CAPTCHA.

3.5 Algorithm for CAPTCHA Breaking

The CAPTCHA breaking consists of further two stages one is Pre-processing and other is Segmentation:

3.6 Preprocessing CAPTCHA

The algorithm for the preprocessing of CAPTCHA is given below:

Step 1: Read the image from the output of design process.

Step 2: Convert the image to grayscale mode.

Step 3: Compliment the image.

Step 4: Mark pixels 1 above value 125 and 0 to the values lesser to the 125. (Threshold Value)

Step 5: Compliment the image.

Step 6: End

3.7 Segmenting CAPTCHA

The algorithm for the segmentation of the CAPTCHA is given below:

Step 1: Read the image from the output of preprocessing CAPTCHA process.

Step 2: Compliment the image.

Step 3: Remove the objects whose pixel value < 50

Step 4: Measure the properties of the image regions

Step 5: For 1 to N (where N is no of regions propied)

Step 6: Crop each region and label it.

Step 7 Store the each cropped region.

Step 8 Repeat steps 5 to 8 till each region completes

Step 9 End

IV. IMPLEMENTATION

The design of the proposed CAPTCHA is implemented in VB.net and the implementation for the comparative analysis has been done in MATLAB 2010b version. The comparative analysis has been done on the designed CAPTCHA (Designed Set) and the existing CAPTCHAs (Sample Set) collected from websites. In this analysis 50 samples were collected out of which 20 samples of both sample set and designed set have been taken on the basis of length and quality of CAPTCHA images. The length of both sets CAPTCHA images has kept same. Out of twenty samples five samples are of 6 character length , five samples are of 5 character length and ten samples are of 4 character length and then the factors for the security measures of CAPTCHA, numbers of segments created, turnaround time and problem solving accuracy is examined.

4.1 Evaluation Criteria

The factors considered for measuring the effectiveness of the proposed CAPTCHA are:

a) Number of Segments

b) Problem Solving Accuracy

Numbers of segments are the segments created after segmentation process. If numbers of segments created are less than number of characters, then the CAPTCHA is hard to segment and is more secure. The robustness of a text CAPTCHA is typically determined by the strength of its segmentation-resistance mechanism.

Problem Solving Accuracy (P.S.A.) is measured by formula which is given below:

$$P.S.A = \frac{\text{Number of Segments Created}}{\text{Length of CAPTCHA}} \times 100$$

Higher is the P.S.A. less strong is the CAPTCHA.

After implementation, following results have been observed:

Table 1 Numbers of Segments Created, Turnaround Time and P.S.A. of Sample Set and Designed Set

Image No.	Number of Characters of	Numbers of Segments Created of CAPTCHAs		P.S.A. of CAPTCHAs (%)	
		Sample Set	Designed Set	Sample Set	Designed Set
1	6	12	1	200	17
2	6	6	4	100	67
3	6	5	2	83	33
4	6	6	1	100	17
5	6	6	3	100	50
6	5	8	3	160	60
7	5	5	2	100	40
8	5	5	3	100	60
9	5	5	2	100	40
10	5	5	3	100	60
11	4	3	3	75	75
12	4	4	2	100	50
13	4	4	3	100	75
14	4	3	2	75	50
15	4	3	2	75	50
16	4	4	2	100	50
17	4	4	3	100	75
18	4	4	3	100	75
19	4	4	2	100	50
20	4	4	2	100	50

V. RESULTS AND DISCUSSIONS

In order to demonstrate the success of the proposed user friendly CAPTCHA, average number of segments created, and problem solving accuracy is calculated and result is analyzed. The comparative analysis of number of segments created of sample set and designed set is given in Figure 4:

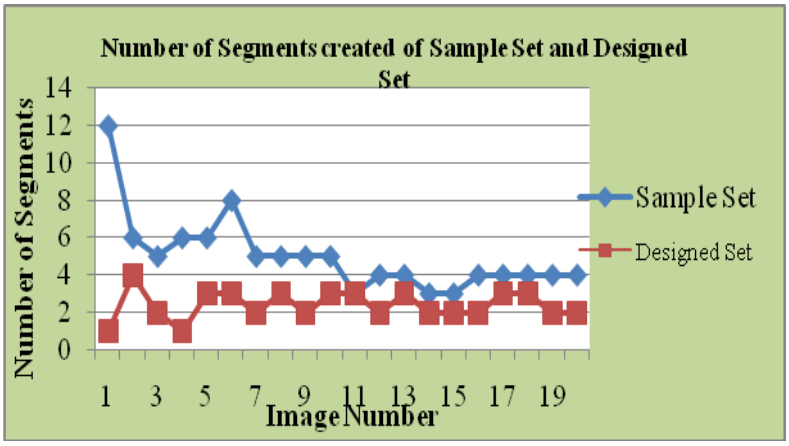

Figure 4 Number of Segments Created of Sample Set and Designed Set

From the above graph in Figure 4, it is observed that the sample set CAPTCHA images have more number of segments than designed set which results that designed set is hard to segment and is more secure.

Another factor for security measurement, comparative analysis of Problem Solving Accuracy of both sample set and designed set is given in Figure 5:

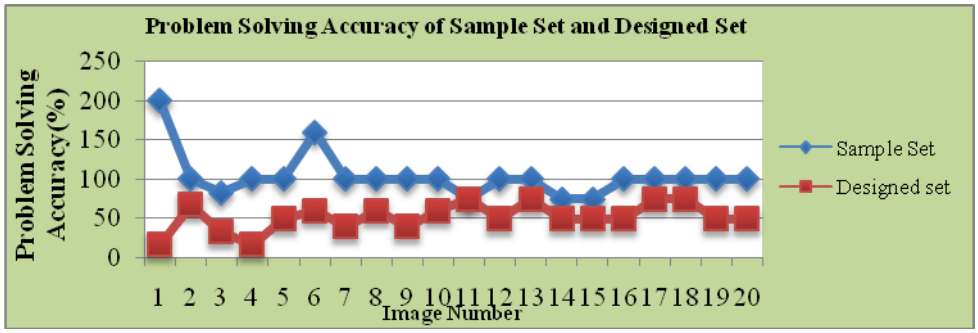

Figure 5 Problem Solving Accuracy of Sample Set and Designed Set

From the above graph in Figure 5, it is observed that the designed set CAPTCHA images have less P.S.A. so is more secure because less is P.S.A. more secure is CAPTCHA. The overall result analysis of average number of segments, turnaround time and P.S.A. is given in Table 2:

Table 2 Comparative Analysis of Sample Set and Designed Set

Sr. No.	Characteristics	Sample Set	Designed Set
1	Average numbers of characters in CAPTCHA	4.7	4.7
2	Average Number of Segments (No.)	5	2.4
3	Probable Solving Accuracy (%)	103.4	52.2

The analysis of the results shown above are as follows:

1. Average numbers of segments created after making an attack on designed set is less than that of sample set. It implies that it is more secure.
2. Problem solving Accuracy tells that only 52.2% characters are segmented of the designed set and 103.4% characters are segmented of sample set. Less the P.S.A. more secure is the CAPTCHA.
3. The proposed CAPTCHA uses only bold uppercase English letters and blank spaces are used between each character which implies that the proposed CAPTCHA is more user friendly. Average numbers of segments for the proposed CAPTCHA is less than the average number of characters of actual CAPTCHA image which implies that it is more secure.

VI. CONCLUSION

In this paper, a user friendly text based CAPTCHA is proposed for secure online authentication. The proposed CAPTCHA is not only user friendly but also secure than existing CAPTCHAs. The proposed CAPTCHA is user friendly because only bold uppercase English letters are used and spacing has been kept between each character of CAPTCHA image so that it is easy for human to read. For the security measurement of the proposed CAPTCHA, a comparative analysis has been done of existing CAPTCHAs and the proposed CAPTCHAs. After the comparison, it was concluded that the proposed CAPTCHA is more secure than existing CAPTCHA because it has resistance over segment creation and has less problem solving accuracy. If problem solving accuracy is less, more secure is CAPTCHA.

REFERENCES

[1] Prof.A.A.Chandavale, Dr. A.M.Sapkal & Dr.R.M.Jalnekar, "A frame-work to analyze the security of text based CAPTCHA", [2010], International journal of Computer Applications (0975 - 8887) Volume 1 Issue 27, ISBN No. 978-93-80746-26-5 pp 127-132.

[2] Suliman A. Alsuhibany, "Optimising CAPTCHA Generation", [2011], Sixth International Conference on Availability, Reliability and Security of IEEE Computer Society, ISBN No. 978-0-7695-4485-4, pp. 740-745.

[3] Elie Burztein, Matthieu Martin, John C. Mitchell, "Text-based CAPTCHA Strengths and Weaknesses", [2011], 18th ACM Conference on Computer and Communication security, ISBN No. 978-1-4503-0948-6, pp. 125-138

[4] Kuo-Feng Hwang, Cian-Cih Huang, Geeng-Neng You, "A Spelling Based CAPTCHA System By Using Click", [2012], International Symposium on Biometrics and Security Technologies, ISBN No. 978-0-7695-4696-4, pp. 1-8.

[5] Wei-Bin Lee, Che-Wei Fan, Kevin Ho, Chyi-Ren Dow, "A CAPTCHA with Tips Related to Alphabets Upper or Lower Case", [2012], Seventh International Conference on Broadband, Wireless Computing, Communication and Applications of IEEE Computer Society, ISBN No. 978-0-7695-4842-5, pp. 458-461.

[6] Xiao Ling-Zi, ZHANG Yi-Chun, "A Case Study of Text-Based CAPTCHA Attacks", Knowledge Discover of IEEE Computer Society, ISBN No. 978-0-7695-4810-4, pp. 121-124.

[7] A. K. B. Karunathilake, B. M. D. Balasuriya and R. G. Ragel, "User Friendly Line CAPTCHAs"

[8] Sushama Kulkarni, Dr. H. S. Fadewar, "CAPTCHA Based Web Security: An Overview", [2013], International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11.

[9] Achint O. Thomas, Sulabh Choudhury, Venu Govindaraju, "Leveraging the Mixed-Text Segmentation Problem to Design Secure Handwritten CAPTCHAs", [2010], 12th International Conference on Frontiers in Handwriting Recognition of IEEE Computer Society, ISBN No. 978-0-7695-4221-8, pp. 13-18.

[10] Kumar Chellapilla Patrice Y. Simard in L. K.Saul, Y. Weiss, and L. Bottou, editors, **"**Using Machine Learning to Break Visual Human Interaction Proofs (HIPs) **",** [2005], Advances in Neural Information Processing Systems 17, pp. 265–272. MIT Press, Cambridge, MA.

[11] Prof. (Mrs) A.A.Chandaywale & Prof. Dr. A.M.Sapkal , **"**Algorithm for secured online authentication using CAPTCHA.", [2010], In proceedings of 3rd International conference on Emerging Trends in Engineering & Technology. pp 292-97, IEEE Computer Society.

[12] A. Caine and U. Hengartner, "The AI Hardness of CAPTCHAs does not imply Robust Network Security", [2007], In Trust Management of International federation of Information processing Publications, ISBN No. 978-0-387- 73654-9 pp 367–382.

[13] Yan, J. and Ahmad, "Usability of CAPTCHAs or usability issues in CAPTCHA design", [2008], In Proceedings of the 4th symposium on Usable privacy and security, volume 337 of ACM International Conference Proceeding Series. ISBN No. 978-1-60558-276-4, pp 44-52.

[14] Kanika Singhal, R S Chadha, "CAPTCHA Generation for Secure Web Services" , [2013], in the proceedings of International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 10, pp 168-170.

[15] Pope, C. and Kaur, K., "Is it human or computer? Defending e-commerce with CAPTCHAs", [2005], IT Professional Journal Published by IEEE Computer Society , Volume 7 Issue 2, pp 43-49.

[16] Prof. (Mrs) A.A.Chandaywale & Prof. Dr. A.M.Sapkal , **"**Algorithm for secured online authentication using CAPTCHA.", [2010], In proceedings of 3rd International conference on Emerging Trends in Engineering & Technology, pp 292-97, IEEE Computer Society.

[17] Rituraj Soni & Devendra Tiwari, "Improved CAPTCHA Method" , [2010], published in International Journal of Computer Applications (0975 - 8887)Volume 1 – No. 25 pp. 92-94.

[18] Kumar Chellapilla and Patrice Y. Simard,"Using Machine Learning to Break Visual Human Interaction (HIPs)", [2004], 17th Conference of Advances in Neural Information Processing Systems, pp. 265-272.

[19] Henry S. Baird and Michael A. Moll and Sui-Yu Wan, "Scatter Type: A Legible but Hard-to-Segment CAPTCHA", [2005], 8th International Conference on Document Analysis and Recognition (ICDAR'05) of IEEE Computer Society, ISBN No. 0-7695-2420-6, pp. 935-939.