 ISSN: 2321-8169
15 – 21

CAPTCHA Techniques: An Overview

Prakash S. Bodkhe
Associate Professor,
Department of Computer Science
G.S.Sci. , Arts & Com. , College, Khamgaon
Dist : Buldana,(M.S.), India
Email-id : psbodkhe@gmail.com

Dr. P. N. Mulkalwar
Associate Professor & Head
Department of Computer Science
Amlokchand Mahavidyalaya,yeotmal
Dist : Yeotmal , (M.S.), India
Email-id : pnm_amv@rediffmail.com

Abstract – Nowadays access of information on Internet has became the need of everyone. Internet offers number of online services such as creating free email accounts, online polling, payment gateways, online banking and many other online commercial and social activities. However these services are often threatened by attacks from automated computer programs called bots. To get rid of from these bot attacks, CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Human Apart) was developed to make the difference between bots and human being. CAPTCHA is a mechanism that offers required security and prevents web services from bots.In this paper we studied most of the CAPTCHA types with respect to their classification, working, breaking procedure, advantages, drawbacks, performance and applications.

Keywords*: Bots, CAPTCHA, web services, Commercial and Social activities, to get rid of*

*****

I. INTRODUCTION

In the modern time human living has been highly influenced by the Internet and its services. It offers number of online services such as education, banking, shopping, communications etc. Most of these services requires permission through filling of online registration form so as to allow the entry to website or specific web page. However, spammers and hackers have developed many unauthentic automated programs called bots which automatically completes the required online registration. The fact is that, the Internet users are not always the legitimate users and sometimes it may be abused by bots. The valuable space is exhausted by these illegal bot programs. To allow access only to the authorized human being, various security systems was developed and deployed.The most common is the password authentication system. Passwords comprises of text, are being used commonly for determining authenticity of users. CAPTCHA is another scheme to test for user authentication. It is a kind of Human Interactive Proof (HIP).

Moreover, CAPCHA is challenge-response type of test which human can solve rather easily but difficult for spammers or bots to pass it.

Before getting online access to web services, a user needs to pass a test provided by password authentication system or CAPTCHA authentication system. The basic problem of using password system is that the login password can be hacked illegally by malicious programs. This invader program can fetch the important data without kind permission by pretending to be a human while login. To address this problem, CAPTCHA was developed to protect the web information from unauthorized users. It is a kind of a program implemented to make the difference between legitimate human user and illegal computer bot program by asking user to type text which is sufficiently distorted or identify image or recognize sound, or solve the puzzle etc. Such tests can be easily passed or solved by human only.

There are number of CAPTCHA techniques which are being used currently by different websites for protecting their information can be broadly grouped into following two categories:

1. OCR-based CAPTCHAs and
2. Non-OCR based CAPTCHAs

OCR-based CAPTCHAs are generally based on text in that user needs to type in text which is distorted sufficiently enough, in the text box.

Non-OCR based CAPTCHAs are generally based on images, audio, video, animation and logics to test the sense of human.

II. Brief History of CAPTCHA

The first use of anything similar to CAPTCHA was back in 1997, when Alta-Vista, a search platform sought a way of blocking automated URL submissions to their search engine. Alta Vista's chief scientist, Andrei Broder developed an algorithm that randomly generated an image of printed text, the first ever instance of CAPTCHA technique. The algorithm was refined by researchers at Carnegie Mellon University in the year 2000, who then termed the technology called CAPTCHA. It is short for Completely Automated Public Turing Test to Tell Computers and Humans Apart. Anyway, computers were unable to recognize it, but for human it was rather easy to read the same and typing it in given text box. This new technology then started adopting by most of the web service providers. A patent was issued to Andrei Broder and his team in April, 2001. Since then there is continuous and rigorous work is being done in this area and still there is a continuous need of new technology or improvements in the existing techniques.

 ISSN: 2321-8169
15 – 21

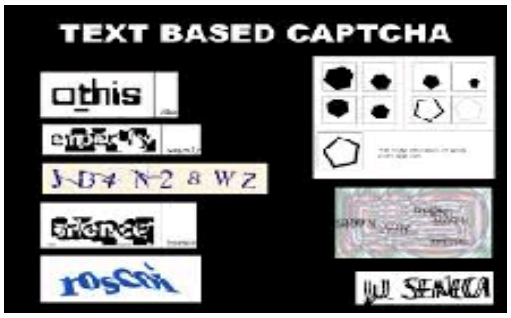
III. Types of CAPTCHAs

CAPTCHAs basically divided on the basis of their design principle and material used for their generation such as text, image, audio, video and some kind of puzzle are:

A. Text-based CAPTCHAs
B. Image-based CAPTCHAs
C. Audio-based CAPTCHAs
D. Video-based CAPTCHAs and
E. Puzzle-based CAPTCHAs

A. Text-based CAPTCHAs :

CAPTCHAs based on text are very easy to implement where text image consisting of alphabetic or alphanumeric characters is presented to the user with distorted image by adding noise, waved characters,3D characters, rotated characters, scattered characters within the text box to obfuscate the user other than human.Then the user is expected to enter the correct text in the text box to pass the test. The following figure illustrates the examples of text-based CAPTCHAs:



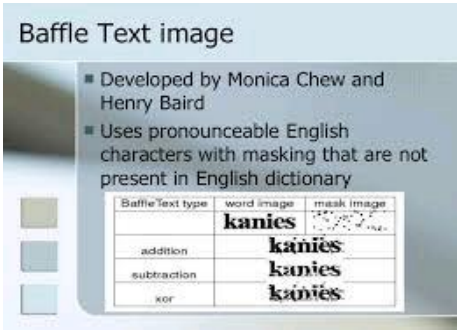


Figure 1 : Samples of text-based CAPTCHA

B. Image-based CAPTCHAs :

Image-based CAPTCHAs are designed by using various image objects. The user has to recognize a specific image to pass the test. Sometime the images are provided with tags and user is asked to identify correct image and enter appropriate word in the box given or asked to click on specific image to prove as a human user. The advantage of image based CAPTCHA is that the pattern recognition is very hard AI problem and thus it becomes difficult to break this test using pattern recognition technique.

Figure 2: Samples of Image-based CAPTCHA

C. Audio-based CAPTCHAs :

ECO was the first Audio–based CAPTCHA developed by Nancy Chan as an alternative to text-based CAPTCHA, as a student at City University, Hong Kong, particularly for visually disabled human users. It uses downloadable audio clips which the user should listen carefully and then submit the listened word in the text box. Audio CAPTCHAs are comparatively more difficult to solve if the sound quality of system is poor.

 ISSN: 2321-8169
15 – 21

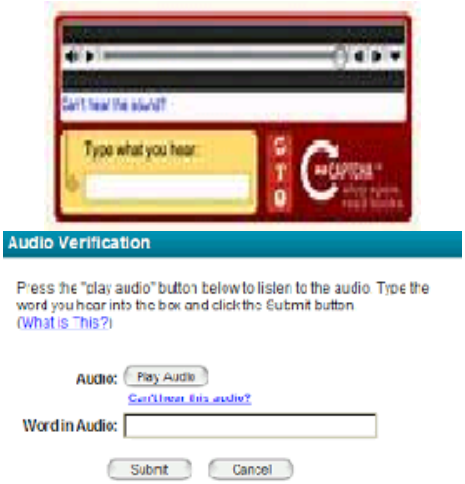


Figure 3: Samples of Audio-based CAPTCHA

D. Video-based CAPTCHAs

NuCaptcha is the world's first video based CAPTCHA that uses motion video to authenticate human web interactions. It is one of the best solution compared to image-based CAPTCHAs, because it is displayed video in video format which is harder-to-recognize for bots and can be read easier by humans which is great. The service is very easy to use and absolutely free of cost. With NuCaptcha basic, you can choose one of the several themes that best fits your performance or website. Even you can customized your captcha skin, background and message displayed in Flash videos which helps a better visual integration with the websites.

Figure 4 : Samples of Video-based CAPTCHA

E. Puzzle-based CAPTCHAs

Generally a puzzle-based CAPTCHA can either a graphics based puzzle or a mathematical puzzle [14]. In picture based puzzle, the picture is divided into some pieces and provide these pieces randomly. Each piece of picture will have piece number. The user needs to arrange these pieces properly by following piece numbers to form a complete original picture [5], the puzzle based on mathematics is hundred percent effective and it can be incorporated into login process and online form registration in the websites for ensuring legitimate access. The user needs to solve that math puzzle provided in order to get legal permission to access web contents.

Figure 5: Samples of Puzzle-based Captcha

IV. RELATED WORK

A lot of research has been made on CAPTCHAs by researchers. There are number of CAPTCHA techniques available to defend the illegal access of important online information by malicious automated bots. The literature review of most of the CAPTCHA techniques is summarized below:

Kunika Singhal [3] has designed robust and user-friendly CAPTCHA. It is generated using Markova text and time variance. The addition of noise with misalignment of characters is done in order to enhance the strength of CAPTCHA.

Prerna Sharma, et al [4] examined all types of CAPTCHAs and in their research they proposed additional layer of security to strengthen the CAPTCHA. They have divide their research work based on identified possible attacks into three main steps:

1. Breaching client-side trust,
2. Manipulating server side implementation and
3. Attacking the CAPTCHA image

They have also suggested the guidelines to exploit the above mentioned vulnerabilities.

Yamanmoto. T. et al [7] in their research on captcha , proposed SS-CAPTCHA (Strangeness in Sentence CAPTCHA). The technique is based on human ability to make the difference in natural sentences composed by human being and computer-translated sentences. In this scheme, user is shown with multiple sentences combining both natural sentences and machine translated sentences. Then, user is expected to select the correct sentence which is composed by a human to prove that he is human and not a bot.

Ved Prakash Singh and Preet Pal [5] in their review on various types of CAPTCHAs, have studied and listed their applications and drawbacks. They mentioned various applications of CAPTCHA on web to protect them from bots including: online polls, web crawling, registration

 ISSN: 2321-8169
15 – 21

of online forms, E-ticketing, Email Spam. They have also listed drawbacks of all types of CAPTCHAS.
The text based CAPTCHA have drawbacks that in text image, user has some difficulty in recognizing the correct text or characters related with more than one font, their size, waved and blurred characters. A well advanced OCR can easily identify such kind of text CAPTCHA. In image-based captcha , some users with low vision or due to blurred images , may find difficulty in image identification. In their work , for the audio-based CAPTCHA , they workout the problem using this kind of CAPTCHA is that it available in English Language only and therefore user is supposed to be very fluent in English. Another drawback of this system is with the characters having nearly similar sound. In the video-based CAPTCHA, the drawback is related with large size of file, and the problem in downloading video. The drawback of puzzle-based is noted by them is delay in solving puzzle and sometimes the arrangement of puzzle.

Sushma Kulkarni and Dr. Fadewar , [2] carried out efforts on Artificial Intelligence (AI) problems that can be implemented for designing the more efficient CAPTCHA that enhanced the security of web contents.

M.Shirali-Shahreza and S. Shirali-Shahreza [10], proposed and designed advanced Collage CAPTCHA to reduce the attacks by providing more resistance to the earlier version of Collage CAPTCHA. In this method, after having selected one of the image among the other image objects, the system then selects other objects by twisting them slightly and set them on the right side of the screen. It uses one additional image which is same as the target image .The user then has to click on the goal image and also on the same image which was placed on the right side of the screen.

Aditya Raj, et al, [1] introduced the novel concept of scrambled CAPTCHA, which combines both the OCR and Non-based techniques. It exploits the inherent characteristics of human visibility and perception. In this paper, they proposed design of Scrambled CAPTCHA. The benefit of this CAPTCHA is that it renders the segmentation attacks ineffective. They have suggested transformation operations of translation, rotation and scaling, shearing and overlapping between characters are applied to each character in a text image. These efforts are able to counter the random-guessing attacks on OCR-based part of it. The fragmentation of pixels in the characters which means that character is not completely colored prevents from CFS (Color-Filling-Segmentation) attacks.

M.Shirali-Shahreza et al. [9] enhance the technique of Motion CAPTCHA by providing user with short movie clip in which a person performing some kind of action. Then some sentences explaining that action is shown is shown to the user. The user is then expected to select the correct explanation. If the user is able to select the correct answer, then he/she will allow to access the web services otherwise system will provide with new test to the user as a second chance. This method is simple and easy to pass but requires good fluency in English language.

Samruddhi , et al. [6] in their paper on different types of CAPTCHA techniques based on drag and drop mouse action gives comparative study of different CAPTCHA techniques. They proposed Drag and Drop (DnD) CAPTCHA technique to counter the possible OCR with respect to text reading CAPTCHAs. In this type of CAPTCHA, the challenge is to simply drag and drop block characters into their different blank blocks instead of entering a textual characters.

Almazyad, A.S. et al. [8] in their research paper proposed a Multi-Modal CAPTCHA method. They combined text-based and image-based techniques in which hundreds of images of different objects such as vegetables, flowers, fruits, animals, trees and many other images are obtained from search engines such as Google, Bing, Yahoo etc. and this large set of images along with tags or captions is stored in the database. The user is then shown an image with four textual labels and user is expected to write correct answer in the box. Multi-Modal CAPTCHA test is very simple to pass as it uses images that are very simple to identify.

C.J.Hernandez-Castro, et al.[11] in their study of pitfalls in CAPTCHA design and implementation , focused on Math CAPTCHA. This kind of CAPTCHA was tested for five mathematical problems: Differentiation, Polynomial, Arithmetic, polynomial question. The user has to provide the solution to the mathematical problem in order to prove that a user is a human and not any automated bot program.

V. Working of CAPTCHA :

Text based CAPTCHA is a widely used method among the different types of CAPTCHAs. The way the CAPTCHA works is illustrated below:

1. Generate some random value: Some string value is created, randomly generated values are always hard to predict.
2. Generate Image: It is found that images are very hard to identify for computers but are very easy for human to read them. The simple text can be easily read and thus CAPTCHA using such a text can be beaked by bots. To make it unreadable for them, designers uses various techniques to make the text image hard enough to read for computers. The techniques uses characters with different characteristics: such as rotated scattered characters, twisted characters, leaning forward or backward characters, 3D characters, cluttering, zig-zag or waved characters, background with multiple lines with different shapes and curves. The text image can use number of techniques to make CAPTCHA image hard and hard so as it becomes difficult for computers to read but easy enough for human users. There is a dynamic need of new techniques as most of the existing techniques are exploited by crackers.
3. Database: The numerous combinations of strings are stored so as to generate text image randomly and matching the same with user input. The simplest solution to this problem is to use Session Variables.
4. Matching: The next step to above is that the CAPTCHA image is generated and apply some techniques to make the CAPTCHA image unreadable to computer but easy for human to read so as to protect from being misused. The

Users fill up the form along with CAPTCHA text and submit it. Now, we have the following:
All submitted form data.
i. CAPTCHA image (from form), which is input by user.
ii. CAPTCHA image (actual one (original), generated by us), from session variable. Session variable is mainly used as it can maintain stored values across web page requests. Now, we required to preserve these stored values from one page (form page) to another (action page).

iv. If both do match, then it is all right otherwise not, in such a case we can prompt the user with message that the CAPTCHA entered was wrong and submission of form is rejected. You may give another chance with new set of CAPTCHA image and ask them to verify.

VI. CAPTCHA Breaking Process

There are continuous efforts and attempts to beat existing CAPTCHA methods by creating programs that include the following functionality:

i. Image Capturing: This is pre-processing method. CAPTCHA image is obtained either from database or by designing your own, following the principles and guidelines required.

ii. Image Cleaning: Removal of background noises and cluttering.

iii. Image Segmentation: This process involves splitting the image into segmented regions which each contains a single character.

iv. Classification: Recognizing then the character in each segmented region.

The steps i, ii and IV are easy for computers to workout. In the process of breaking CAPTCHA, segmentation is the only step which outperforms computers efforts. If the background clutter have some shapes similar to character shapes, or these are connected by clutter, the segmentation then becomes quite impossible by recent software. Thus, CAPTCHA designers should focus on the segmentation part so as to construct segment-resistant robust CAPTCHA. The following figure shows how OCR is used to clean the corrupted text image to its original form.

Figure 6 : The text based CAPTCHA identified by OCR.

VII. Advantages of Different types of CAPTCHA

There are many advantages related with each type of CAPTCHA. Type wise advantages given below:

a. Advantages of Text-based CAPTCHA :

- It is the most popular and widely adopted CAPTHA type by most of the websites because of its simplicity.
- It user-friendly and comparatively simple to implement.
- Baffle text based CAPTCHA is used to defends the resources from dictionary attacks.[12].
- reCAPTCHA as a text based CAPTCHA which provides service to digitize books ,newspapers and handwritten literature and enable you to design your own CAPTCHA.

b. **Advantages of Image-based CAPTCHA** :

- It is used as an alternative to text-based captcha.
- It uses mouse in some version of image-based CAPTCHA to click over
- Using this CAPTCHA, pattern recognition of image is difficult Artificial Intelligence problem.

c. Advantages of Audio-based CAPTCHA :

- It is useful for blind persons or persons with low vision.
- It is user-friendly. [13]

d. Advantages of Video-based CAPTCHA:

- It is very hard to break by OCR.
- It defends laundry attacks.
- provides more security than text-based and image-based captcha.[12]

e. Advantages of Puzzle-based CAPTCHA:

- It is simply a fun to use.
- It helps to check your mental ability.
- It is something like a game which user can interact with more interest.[13]

VIII. Pitfalls in Different types of CAPTCHA

There are some drawbacks of using CAPTCHA which are listed below. Their use is not always 100 % effective against bots. There are some tools that allow for bots to read and capture the text of images which then makes the CAPTCHA useless.

a. Text-based CAPTCHA issues :

- In this kind of CAPTCHA , user find difficulties in identifying and entering appropriate text or some letter or character due to the reasons such as :different font size, color, orientation, use of blurred characters, multiple fonts, use of various shapes and lines etc.
- Text-based CAPTCHAs are OCR sensitive and may be beaten often by them.
- The blind people or person with low visibility cannot easily pass the test. [12]

b. Image-based CAPTCHA issues :

- Not suitable for people having color blindness.
- It can be identified by: random guessing or applying picture-dictionary attacks.
- Requires large database of thousands of images.
- Take up more resources of system and increases the load over server.
- Problem of misspelling or miss-tagging of images.

c. Audio-based CAPTCHA issues :

- It is not suitable for those user who have poor vocabulary of English. It needs thorough knowledge of English language as it is available only in that language.
- Problem in hearing sound of similar characters.
- Not suitable for deaf users or human with low hearing power.[13]

d. Video-based CAPTCHA issues :

- Sometimes the video file size has becomes an issue while downloading video and then to pass test graded by that video.
- Video speed.[12]

e. Puzzle-based CAPTCHA issues :

- Sometimes difficult for user to identify puzzle test.
- May take longer time to identify and to solve the puzzle.[13]

IX. PERFORMANCE OF CAPTCHA

From the above study, it is now quite possible to measure the performance by comparing all the types of CAPTCHAs with their security issue, and ability of its users. The following table lists the security and usability of different types of CAPTCHAs with respect their user ability [13] :

Captcha Type	Security Provided	Usability	
		Usable for visible disable users	Easy / Difficult to use
Text-based CAPTCHA	Good	No	Average
Image-based CAPTCHA	Good	No	Easy
Audio-based CAPTCHA	Good	Yes	Difficult
Video-based CAPTCHA	Good	No	Difficult
Puzzle-based CAPTCHA	Average	No	Depends on type of user and type of Puzzle test Difficult / Easy

X. Applications of CAPTCHA :

CAPTCHAs have number of applications, some of the important one are illustrated below:

A. Protecting Registration forms in website :

Some websites like Yahoo, Gmail, Hotmail, Facebook, Twitter etc. provides free registration to the users. These websites uses incorporated CAPTCHAs to protect the registration process. Thus, any website which offers free registration, it is always good to set up a CAPTCHA to ensure that all the registrations are done by human user and not by the automated bots.

B. Protecting Email Accounts :

Email accounts of many people are hacked or their Email accounts have been taken over by hackers and spammers and you cannot login to your account anymore. The website then may be misused by the spammers or hackers. To address such an issue, Email service provider websites such as Yahoo, Gmail, Hotmail, Rediffmail etc. required their users to go through a CAPTCHA test whenever they want to use.

C. Protecting Online Shopping :

In order to authenticate that the buyer is a person and not any automated bots, the online shopping websites generally implements CAPTCHAs to protect their products from the fake orders.

D. Protecting worms and spam comments :

Some social websites such as Facebook or Twitter which invites comments from public on some specific issue are bombarded with hundreds of comments with single post. The search engines such as Google, Yahoo, and Bing etc. would assume that the websites having multiple comments on same single post are inviting spams. To overcome this crucial issue, it is necessary for such sites to use CAPTCHA, so that only human users will post comments and not the spammers.

E. Protecting online polls

The online polling is conducted on many social and governmental websites for various purposes. If the voting is done illegally by the automated bot programs or spammers, then there will be a question mark on results and conclusions made on the conducted polls. Thus, these websites conducting online polling should ensure that no one but only human can vote by using CAPTCHAs.

F. Phishing Attacks

Phishing is the attempt to obtain sensitive information such as usernames, passwords, and credit card details (and, indirectly money), often for malicious reasons. Fishing is an email that falsely claims to be a legitimate user in an attempt to scam the user into surrendering private information or login details. Use of CAPTCHA is the prevented treatment to defend such malicious attempts by scammers.

G. Protecting Search Engine Bots

It is generally expected to keep web pages without index to prevent others from getting them easily. An html tag is used to prevent search engine bots from reading web pages. The tag, in fact, don't ensured that bots will not read a web page , it only prompt you to say "No Bots , Please !". Therefore, in order to ensure that bots will not intrude a website, CAPTCHAs are required.

XI. CONCLUSION AND FUTURE SCOPE

CAPTCHAs plays vital role in web security where they prevents hackers, spammers, automated bot programs from misusing online web services. This paper has presented concepts, brief history of CAPTCHA, types of CAPTCHA, their working principle, CAPTCHA breaking process.

In addition we also discussed applications, advantages and drawbacks of each type of CAPTCHA. Finally we compared all types of CAPTCHA to measure their performance based on security and its usability. Therefore, there is a never ending need to develop stronger and more robust CAPTCHA that should always be easy for human to solve but difficult for malicious bot programs.

Future Work

After going through number of research papers describing different types of CAPTCHA, we noticed that all these CAPTCHA designs are relied on English Script. Many Indian regional websites content is given in regional languages besides English language and these websites use CAPTCHA tests in English language in regional language web page. This reduces usability and accessibility because non-native speakers of English are required to pass CAPTCHA tests in English language.

The accessibility of such websites can be improved if CAPTCHA test is provided in regional language. Therefore it is utmost need to design CAPTCHA in regional language scripts such as Hindi, Marathi,Gujrathi, Nepali,Marwari,Bhojpuri etc have many common vowels and cosonants , are spoken by millions of people and collectively known as Devanagari.

Thus, in future, main focus will be on to provide the CAPTCHA in Devanagari script .

REFERENCES

[1] Aditya Raj, Tushar Pahwa , Ashish Jain [2013], "Countering Spam Robots : Scrambled Captcha and Hindi Captcha ", DOI : 10.4018/978-1-4666-078- 5.ch019.

[2] Shushma Kulkarni, Dr. H.S.Fadewar, " CAPTCHA Based Web security : An Overview ",[2013], International Journal of advanced Research in Computer Science and Software Engineering , vol. 3,Issue 11.

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

[4] Prerna Sharma, Nidhi Tyagi and Deepali Singhal, "CAPTCHA : Vulnerability To Attacks " , 2013, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), vol.2, Issue 2, March-April 2013.

[5] Ved Prakash Singh, Preet Pal,"Survey of Different types of CAPTCHA", International Journal of Computer science and Information Technologies, vol. 5(2),2014, pp. 2242-2245.

[6] Bhalani, Samruddhi D., & Shailendra Mishra,(2015),"A survey on CAPTCHA Techniques Based on Drag and Drop Mouse Action", International Journal of Technical Research and Applications, vol.3 Issue 2, pp. 188- 189.

[7] Yamanmoto T., Tygor J.D. and Nishigaki,M.(2010), "CAPTCHA Using Strangeness in Machine Translation", Advanced Information Networking and Applications(AINA), 2010,24th IEEE International Conference on, Perth, WA,2010,vol.no.pp. 430-437.

[8] Almazzyad, A.S., Ahmad. Y. & Kouchay, S.A.(2011), "Multi-Modal CAPTCHA : Auser verification Scheme ", Information Science and Applications (ICISA), 2011, International conference on., Jeju Island,2011, vol.no.pp.1-7.

[9] Shirali-Shahreza, M & Shirali-Shahreza S. (2008),"Motion Captcha", Human System Interactions, 2008, conferenceon., KraKow,2008,vol. no. pp 142-144.

[10] Shirali-Shahreza, M & Shirali-Shahreza S. (2008), "Advanced Collage CAPTCHA", Information Technology : New Generation, 2008, ITNG , Fifth International Conference on., Las Vegas, NV 2008, vol no. pp. 1234-1235.

[11] C.J.Hernandez-Castro, A. Ribagorda, "Pitfalls in CAPTCHA Design and Implementation: The Math Captcha , a case study", Computers and Security, 29(1),pp.141-157,2010.

[12] Walid Khalifa Abdullah Hasan, , "A Survey Of Current Research On CAPTCHA ", International Journal of Computer Science & Engineering Survey (IJCSES) vol. 7, no.3 June 2016.

[13] Kiranjot Kaur and Sunny Behal, "Captcha and Its Techniques : A Review ", International journal of Computer Science and Information Technologies. 2014, vol. 5(5), pp. 6341-6344

[14] Divyashri N. and Dr. T. Satish Kumar , " A Survey on Captcha Categories ", International Journal of Engineering And Computer Science ISSN : 2319-7242, volume 5 May 2016, page no. 16458-16462.