

# A Novel CAPTCHA Design Approach using Boolean Algebra

Ramanpreet Kaur  
M. Tech Scholar (CSE)  
CTIEMT, Shahpur Jalandhar

Pooja Choudhary  
Assistant Professor (CSE)  
CTIEMT, Shahpur Jalandhar

## ABSTRACT

CAPTCHAs are employed on web systems for strengthen web security to protect from internet bots, fraudulent registration, spammers. So CAPTCHA is designed in such a way the patterns display on CAPTCHA feature that human can easily pass but hardship for automated approaches. Usability and security are two rudimental issues for CAPTCHA design and implementation. In Math calculus CAPTCHA test that be inherent with trigonometric and differential functions. Due to complexity of equations, user cannot easily solve expressions that possess poor usability of CAPTCHA. In this paper, a novel CAPTCHA design approach using Boolean algebra is present, to improve usability and security as contrast to Math calculus CAPTCHA. In this approach, logic gates will be design that depict shapes like OR, AND, NOR, NAND, NOR, etc. by use online digital library tool. Each gate executes a Boolean expression as represent by shape indicating operation. The shapes relate to the logic gates like conjunction, disjunction and complement. To prove human, user will identify shape of given logic gates and choose correct option that display in form of Boolean algebra for pass CAPTCHA test. From security issue, Boolean algebra CAPTCHA is non OCR based that provide no development in advance technologies like AI and automated software.

## General Terms

Security, Usability, Random image function

## Keywords

CAPTCHA, Boolean algebra, Math CAPTCHA, usability, security, bots

## 1. INTRODUCTION

Due to the vast growth of internet, security becomes crucial issue for web applications. Nowadays a large number of people use internet for several activities like education, e-transaction, entertainment, etc. There are many economical reasons to misuse these online services like email account creation, misuse of remote voting system, adding links for commercial promotion, etc. So protect from automatic abuse system, Completely Automated Turing test to tell Computer and Human Apart or CAPTCHA employ on different web applications as a security mechanism to make distinction between human and robots by allow user to pass the CAPTCHA challenge test. CAPTCHA is used as a guard system to keep safe web system from the harms of attackers, malicious bots, spammers, worms, Trojans, virus, fraudulent registers. Many CAPTCHAs have been designed and run on different websites like Google, Yahoo, Microsoft's MSN, Hotmail, and YouTube. These are widely accepted good CAPTCHAs that hold high usability and security. A good CAPTCHA having three desirable properties that are:

- a. **Automated:** Tests should be simple to be automatically generated and graded by a computer.

- b. **Usable:** Usability of system is how it easy for human. Test should easily pass by user in reasonable amount of time.
- c. **Secure:** Test should be difficult for automated devices or OCR to solve.

## 1.1 Origin of CAPTCHA

In 1996, Moni Naor was first person who summarizes a list of methods to identify that a request comes from a user and not a robot [15]. In 1997, the first use of CAPTCHA test was done in Alta-Vista web search engine [1]. The problem seem to complications of automatic submission of URLs. Primitive CAPTCHAs have been developed by Andrei Broder, Martin Abadi, and Mark Lillibridge to obstruct bots which adding URLs to their search engine. In 2000, Luis Von Ahn, Manuel Blum & John Langford at CMU (Carnegie Mellon University) coined term CAPTCHA [1, 15].

## 2. TYPES OF CAPTCHA

The different types of CAPTCHAs have been developed with their strength and weakness that describe below:

### 2.1 Text based CAPTCHA

Text CAPTCHAs are OCR based having alphanumeric characters with distortion and degradation apply on characters [15]. The text based CAPTCHAs are extensively developed and majority used text CAPTCHA system in real world application.

- a. **Gimpy:** It is first visual text CAPTCHA implementation, built by Carnegie Mellon University with Yahoo for their Messenger [17, 19]. That picked randomly English words as rendered them with distortions and shape deformation of letters [15].



Fig 1: Gimpy CAPTCHA [17]

- b. **EZ-Gimpy:** It is simplified version of Gimpy CAPTCHA i.e. EZ-Gimpy, adopted by Yahoo [17, 19]. This CAPTCHA has more character distortion as compared to Gimpy but it is weak due to the ease of segmentation.



Fig 2: EZ-Gimpy CAPTCHA [17]

- c. **Baffle text:** It is a variation of EZ-Gimpy. This CAPTCHA is simple OCR based CAPTCHA that select an English words instead of dictionary word having white-black background [2].



Fig 2: Baffle text [2]

- d. **Google:** This CAPTCHA uses display a random number of lower case alphabets. Due to close character of packing, it is hard to segment [14].



Fig 4: Google CAPTCHA [14]

- e. **MSN:** Microsoft CAPTCHA based on complex OCR uses 8 characters uppercase & digits. This CAPTCHA has a large space of characters to choose from and is hard to segment due to noise [19].

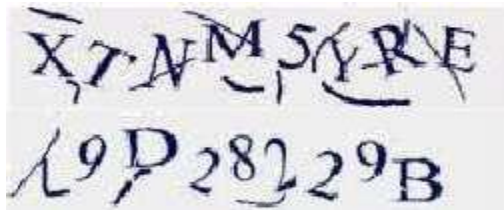


Fig 3: MSN CAPTCHA [19]

## 2.2 Image recognition CAPTCHAs

The current OCR based CAPTCHA become insecure and exploited. For alternative approach of OCR CAPTCHA, the non OCR based CAPTCHA is used that include pictures, audio, video, etc. General visions seem to images are difficult to break by machines than character recognition, so more focus done on pictures instead of character recognition [6]. In graphic CAPTCHA challenge test that involve objects or images in which user have to identify the display image or to guess those images which have some similarity.

The basic images CAPTCHAs are ESP PIX in which different images are shown and user has to select a description from the predefined images [8, 21]. Warner created kitten Auth CAPTCHA in which some variant of images of cats that nine pictures are cute animals and three of which are feline. The problem is database of pictures are small enough (<100) [15]. Microsoft developed ASIRRA CAPTCHA for overcome the kitten Auth. It uses giant database composed of pets and cats and ask the user to select cats in it [15,23].



Fig 4: ESP PIX [21]



Fig 5: Kitten AUTH [22]

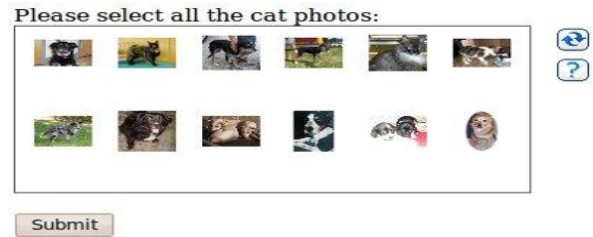


Fig 6: ASIRRA [23]

The images CAPTCHAs show high usability and low security. The availability of short database cause a small number of possible solutions by which random number of guessing can have a high probability of success. For defeat this problems, number of CAPTCHAs designed that embed on composite multiple images apply on tags and geometric transformation. SEMICHA and tagging based CAPTCHA is one of them that based on pre-tagged images using upright orientations or geometric transformation to enhance security [5].



Fig 7: a) Input b) final 3D image [5]

## 2.3 Audio based CAPTCHA

The audio or sound based ECO system was implemented by Nany Chan a student from City University in Hong Kong [2]. This CAPTCHA is mainly used by visually disable person. It shows low usability and high security. For solve the CAPTCHA, user first download audio clip, after listen then submit spoken words [1]. The audio clip is available only in English vocabulary. Due to language dependency every user cannot solve this CAPTCHA.

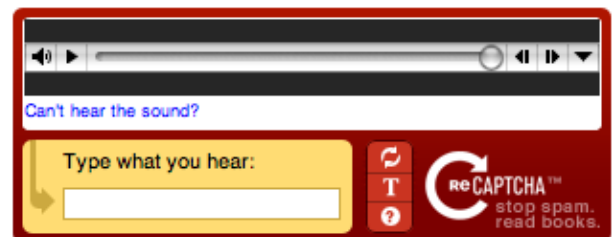


Fig 8: Audio based CAPTCHA [26]

## 2.4 Video based CAPTCHA

Video based CAPTCHA lesser used and difficult to implement. It is very hard for bots to break and more secure than other CAPTCHAs [20]. In this, moving words are displayed on complex background. For pass CAPTCHA user will fill same moving words which show in CAPTCHA image for correct answer. The moving words that are very difficult to break by bots and other automated software because the words have no fixed pixel position. The pixel position may be change in every interval of time. This CAPTCHA is limited use on websites.



Fig 9: Video based CAPTCHA [20]

## 2.5 Logical CAPTCHA

The logical CAPTCHA based on intelligence ability have simple logical questions. For this service database consist of millions of questions [24]. The question is such as:

1. Which of 45,23,54,67,87,98,76 is greatest?
2. Which letter repeated maximum time in word "Maharashtra"?

Table 1. Comparisons of Different Types of CAPTCHAs

S. no	Different types of CAPTCHA	Limitations
1	Text based CAPTCHA	This CAPTCHA assures less usability and security. Due to complex structure of characters like multiple fonts, blurred letters, multiple fonts, etc. the user cannot easily use this CAPTCHA. It can easily identified by OCR strategies.
2	Images based CAPTCHA	Mainly 10 to 12 images are used corresponding to single image CAPTCHA that increases space complexity. The small database cause brute force attack on CAPTCHA. If database is hacked due to less security, it requires more screen spaces than text CAPTCHA. Visually disabled person cannot easily identify images due to blurring and low vision of images.
3	Audio based CAPTCHA	Audio clip only in English vocabulary. Due to language dependency every user cannot use this CAPTCHA. Characters have similar sound and background noise.
4	Logical based CAPTCHA	The questions are specified only in English language. For solve the answer, need intelligence ability. AI devices can easily break this CAPTCHA by machine learning attack. Logical CAPTCHA use real words than random sequence.
5	Video based CAPTCHA	Difficult to implement and time consuming. The user face problem to download the large files and find correct CAPTCHA.

## 3. RELATED WORK

There are several published work about CAPTCHA techniques with their strength and weakness and their issues.

Sarika Choudhary et al. [1] describe CAPTCHA types with their pros & cons, applications and guidelines to construct CAPTCHA. Kiranjot kaur et al. [2] introduce CAPTCHA their types, applications, drawbacks. Durgesh samariya [3] design a new way of CAPTCHA such as splitting image in new parts then characters and digits rotate on different rotational value and add different random lines as background that prohibit segmentation of characters by robot or OCR and thus helps to make more secure CAPTCHA. Shanker, Prashant Gupta, Aditya Jaiswal [4] improves resistance of Collage CAPTCHA by Hybrid Collage CAPTCHA by showing corresponding images with distorted name regarding goal image on right side. In this, user can easily identify the goal image due to small set of pictures given. It shows high usability and response time as comprise to collage CAPTCHA. Maryam Mehrnejad, Ehsan Toreini, Ahad Harati [5] focus on remove problem of limited database of image based CAPTCHA. To improve this problem, two different approaches are used. The first one is called Tagging image CAPTCHA based on pre tagged images using geometric transformations. The second one is SEIMCHA uses apply on upright orientation. Both techniques used for increasing security & usability. In future, eliminate the problems of Tagging CAPTCHA makes more securable and usable. Jing-Song Cui [6] proposed a CAPTCHA in the form of 3D animation that prevent attacks based on both image based and moving object recognition in videos. In this zero knowledge per frame principle is used which ensures that each frame of animation would not lose any information about verification code's content. So it makes difficulties for OCR techniques to attacks on CAPTCHAs but this method have still defects and not a practical used. Renu Dhir, Mandeep kumar [7] proposed a new color based CAPTCHA which provides color based images likes flower, cars, monitor etc. to human and human will answer to interrogator with color name. These colored images have single or more than one color. This CAPTCHA is simple to use, user friendly, security improved but failure for color blindness people and cannot easily recognize the colors. Yogdhar Pandey et al. [9] proposed scheme clickspell that combines the features of both text and image based CAPTCHA. To pass click spell test, user have to spell a randomly chosen word by clicking on distorted letters. In addition, Clickspell can add an advertisement image optionally as background to cover its distorted letters and thus letters cannot be easily segmented and recognized by OCR and resilient to computer automated attacks. Elie Bursztein et al. [10] In this using various automated attacks apply on real world CAPTCHAs and evaluate strength and weakness of different CAPTCHAs. Based on this evaluation, identify constraints of design and implementation and in future direction help to make a reliable CAPTCHA. Nitish Payal et al. [12] proposed a drag and drop image based CAPTCHA by integrating image CAPTCHA with JIGSAW and Puzzle for easy access of web services in lesser time. To execute each and every moves on webpage by dragging images to the correct location without reloading page again and again. Ajax technology is used. This method much interest rather than matching and writing complex clatters text. But in this user take a lot of time to identify actual arrangement of puzzles. Jeff Yan et al. [13] Introduce usability issues in different CAPTCHAs system like text CAPTCHA, Color CAPTCHA. Ahmad El Ahmad and Jeff Chan et al. [14] describe use of colors in CAPTCHA that affects on usability and security



issues of CAPTCHA. In usability, user can easily pass color CAPTCHA by identify color in image that distinguish other colors. Same as, in security case, cracker identifies foreground and background color and easily extracts text from background images. Gaurav Goswami et al. [8] Introduce for overcome the limitation of language dependency text based CAPTCHA use face detection color based CAPTCHA. In this user identify the distorted human faces from the complex background image. This CAPTCHA give better usability, accuracy and lower automated attack. BRIAN M. POWELL et al. [11] proposed a novel image-based CAPTCHA that enhance face detection color based CAPTCHA by using touch-based mobile devices with face detection tests that is easy for humans to solve and give a high security. The optimized fgCAPTCHA works efficiently on both touch screens used by tablets and smart phones achieving a high human accuracy. Carlos Javier Hernandez Castro et al. [15] proposed a “Math CAPTCHA” or “QRBGS CAPTCHA” that request user to solve math problem to prove human. Due to complexity of equations that posses low usability, accuracy and response time. This CAPTCHA scene suffers a lack of methodological analysis and design of the schemes.

#### 4. PROBLEM DEFINITION

A good CAPTCHA provides human validation, fair access and perimeter threat protection. In Math calculus CAPTCHA test, CAPTCHA image subsist with calculus functions that can be inherent with trigonometric and differential equations or ask user for prove human to pass this test. An ideal CAPTCHA is that which is effortlessly solve by user in less time but laborious for hacker. Due to complexity of equations, math calculus CAPTCHA takes a lot of time to solve the expression. Such CAPTCHAs are very annoying and ridiculous for user that give poor usability i.e. not easily solve by user. The technical users also distracted from this type of CAPTCHAs. For solve this CAPTCHA, require a theoretical formula memorize and extent experience or skill about the calculus logic.

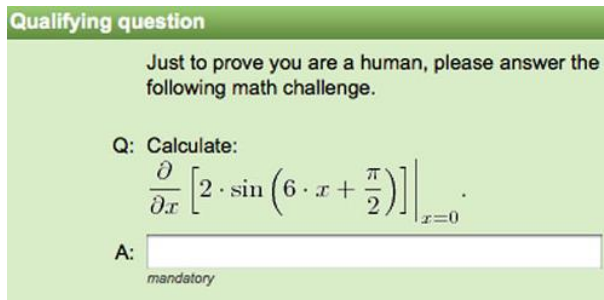


Fig 10: Math calculus CAPTCHA [15]

From security issue, there are various pitfalls in design and implementation. The problems related to the non uniform distribution of all the parameters involved in a CAPTCHA challenge like input parameters, functions chosen, distribution of correct answers, distribution of different challenge characteristics, etc. [15].

$$\frac{\partial}{\partial x} \left[ 3 \cdot \sin \left( 4 \cdot x + \frac{\pi}{2} \right) \right] \sin_{x=0}$$

$$\frac{\partial}{\partial x} [4 \cdot \sin(3 \cdot x)] \sin_{x=\pi}$$

Fig 11: Different Sized Mathematic Symbols

In this, formulas are rendered using symbol that are relatively

positioned each other in normal way as a text with no distortion and overlapping of symbols. OCR strategies easily recognized the normal text by using shape recognition algorithms and pixel or pattern matching algorithm. Shape recognition algorithm work with higher number of pixels like parenthesis, brackets. Pixel matching algorithm using hamming distance detects the similar number of characters. So easily identify symbols by OCR strategies and find correct result by machine learning attack that poses a serious risk to their security.

#### 5. PROPOSED METHODOLOGY

A novel approach Boolean algebra CAPTCHA also called Digital CAPTCHA which will improve resistance of math calculus based CAPTCHA. Boolean algebra is a subarea of algebra that uses to analyze digital gates and circuits. Boolean algebra is a technique of mathematics based on logic reasoning that have set of rules or laws which are used to define and minimize Boolean expressions. In Boolean algebra, an expression given can also be converted into a logic diagram using logical AND, OR, NOT.

To design a CAPTCHA, it will leave out the demerits of Math calculus CAPTCHA. The new CAPTCHA will easy to use, time saving, reduce complexity and same time hard for attacker to recognize a CAPTCHA as contrast to current CAPTCHA.

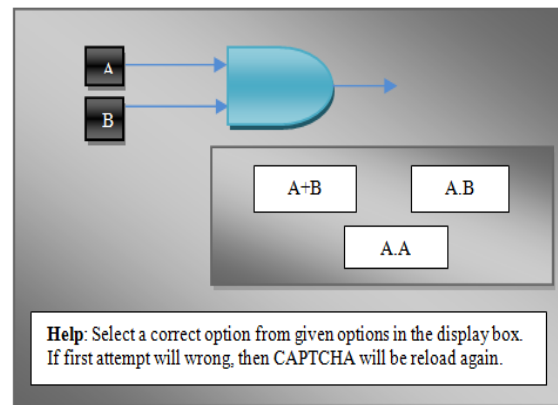


Fig 12: Example of Proposed CAPTCHA

In proposed CAPTCHA scheme, Boolean algebra CAPTCHA construct by using either one or more logic gates by use different input symbols. The shapes relates to logic gates like conjunction, disjunction and complement. Each gate executes a Boolean expression as represented by a shape indicating operation. To improve usability, the options related to Boolean expressions will display. User will select a correct option from given display box to prove human. After each attempt, the CAPTCHA will reload again.

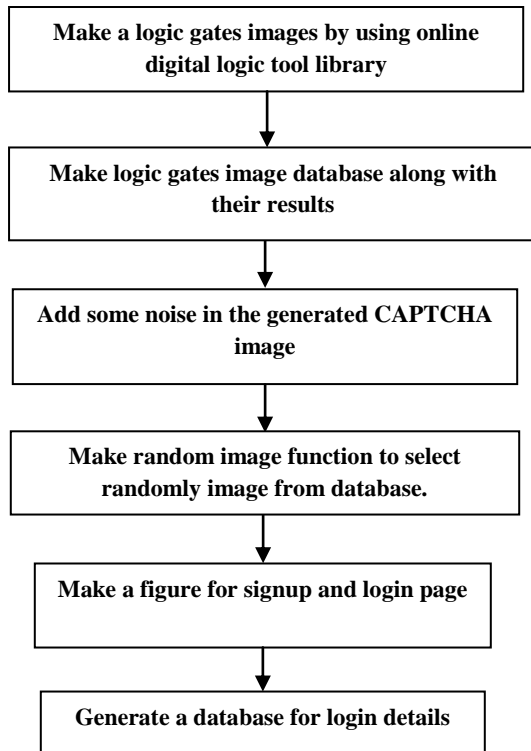


Fig 15: Steps to design a Boolean CAPTCHA

## 5.1 Comparisons and Discussion

To implement Boolean expressions and operations (AND, OR, NOT, NAND, NOR, XOR) which posses high usability, accuracy and response time as contrast to math calculus CAPTCHA. It is simple to evaluate Boolean operation by using law of Boolean algebra. But, in math calculus CAPTCHA i.e. time consuming, calculations take a lot of time for correct result. Boolean algebra is based on logic reasoning, predicate calculus that is easy to understand and memorize as comparison of annoying and ridiculous calculus formulas. To improve usability, pushbuttons will use instead of textbox. It will help in time saving for fill syntactical expression in textbox. By use little knowledge about logic gates, user can use this CAPTCHA.

From security issue, text CAPTCHAs built with alphanumeric codes that can easily scan by OCR by pre-processing, segmentation, recognition. In pre-processing, remove background lines and dot removal. Segmentation process help to remove close character packing. Recognition helps to identify characters and remove noise between characters. Boolean Algebra CAPTCHA consists of lines, pictures i.e. non OCR based. If attacker tries to find shape, pixel position, color by shape and pixel recognition algorithm that is not meaningful that provides no development in advance technologies like AI and automated software. But, in math calculus CAPTCHA has different sizes of same symbols as non trivial way. That can easily recognize by OCR strategies using shape recognition or pixel matching algorithm that lead to high risk to security. In Boolean algebra, use a logic gates that depict different shapes like AND, OR, NOT, NOR, NAND, EX-OR, EX-NOR. The OCR strategies cannot recognize the shape of logic gates, difficult to harder for machines to crack images and also help to prevent from automated attacks.

## 5.2 Advantages

- Boolean algebra CAPTCHA will help to improve

usability and security as contrast to math calculus CAPTCHA which human can easily solve but not by bots.

- Pushbuttons are used instead of text box. It will help in time savy for fill syntactical expression in textbox. By use little knowledge about logic gates, user can use this CAPTCHA.
- This CAPTCHA is non OCR based that lead to no development in artificial intelligence and character recognition software.
- It will overcome the complexity of math calculus CAPTCHA by use simple Boolean algebra laws.
- The logic reasoning based Boolean algebra simple to use by every average person. No need to memorize hard theoretical formulas to solve this CAPTCHA.
- It reduces time complexity. By use simple Boolean laws and solve this CAPTCHA.

## 6. APPLICATIONS

- Comments Spam Prevention:** You have a blog and notice and get a numerous of spam comments. For this CAPTCHA is used to figure out his problem that help to make traffic free websites.
- Registering the web forms:** Free registration services available on different websites. For secure the websites, protect from bots, attackers and only user can access their free services. For this CAPTCHA is necessary to protect all free services from internet.
- Search Engine Bots:** Some confidential pages on your website that you don't want search engine bots to reveal. To prevent all kinds of bots, including those of the search engines, CAPTCHA is used.
- Preventing Dictionary Attack:** CAPTCHAs are also used to prevent dictionary attacks in password systems. To prevent a computer from able to iterate through entire space of passwords by require it to solve a CAPTCHA after a certain number of unsuccessful logins.
- Free Anti Spam:** Mostly companies want to accept trust worthy emails. For this, CAPTCHA use as anti spam filters.
- Online Survey and Polls:** During polling, a candidate wins by get maximum votes. If you wish to get trustworthy result of polling, then use of CAPTCHA is necessary. That will be ensure the polling is done by user not by bots.

## 7. CONCLUSION

CAPTHCHA is widely research field act as internet rectifier to secure web applications by discern human from bots. In this paper, Boolean algebra CAPTCHA presented which will improve resistance of math calculus CAPTCHA. By use, Boolean operations and expressions instead of trigonometric and differential function which will help in reduce the complexity of CAPTCHA and help to achieve better usability and security as compared to math calculus CAPTCHA. Boolean CAPTCHA can be easily use by educated user. No need of technical skill, by using intellectual mind to solve this CAPTCHA and help to reduce time complexity. Due to non OCR based, hard for machines to break this CAPTCHA.

The proposed system is in progress phase. The evaluation of

result and performance on the basis of its result calculation will be implemented as future work.

## 8. ACKNOWLEDGEMENT

The authors would like to thanks to the earlier work regarding different CAPTCHA techniques that contribute the work made in this paper. All work done in this paper will surely help to the researchers for future work on CAPTCHA techniques.

## 9. REFERENCES

- [1] Sarika Choudhary, Ritika Saroha, Yatan Dahiya, Sachin Choudhary, "understanding CAPTCHA: Text and Audio Based Captcha with its Applications", International Journal of Advance Research in Computer Science and software Engineering, Volume 3, Issue 6, June 2013.
- [2] Kiranjot kaur, Sunny Behal, "CAPTCHA and its Techniques: A review", International Journal of Computer Science and Information Technologies, <http://www.ijcsit.com/docs/Volume%205/vol5issue05/ijcsit2014050576.pdf>.
- [3] Durgesh Samariya, "Designing CAPTCHA for Web Security Using Rotating and Spilting Image against OCR Software", International Journal of Advance Research in Computer Science and Management Studies, Volume 3, Issue 2, February 2015.
- [4] Divya Shankar, Prashant Gupta, Aditya Jaiswal, "Hybrid Collage CAPTCHA", International journal of scientific & engineering research, Volume 4, Issue 1, January-2013.
- [5] Maryam Mehrnejd, Abbas Ghaemi Bafghi, Ahad Harati, Ehsan Toreini, "SEMICHA: A New Semantic Image CAPTCHA Using Geometric Transformations," The ISC Int'l Journal of Information Security, September 2012, volume 4, number 1(pp, 63-76).
- [6] Jing-Song Cui, Jing-Ting Mei, Wu-Zhou Zhang, Xia Wang, Da Zhang, "A CAPTCHA implementation Based on Moving Objects Recognition Problems", 2010 International Conference on E-Business and E-Government.
- [7] Mandeep Kumar, Renu Dhir, "Design and Comparison of Advanced Color based image CAPTCHAs", International Journal of Computer Applications, volume 61 – No.15, January 2013.
- [8] Gaurav Goswami, Brain M. Powell, Mayank Vatsa, Richa Singh, Afzel Noore, "FaceDCAPTCHA: Face detection based color image CAPTCHA", ELSEVIER Future Generation Computer Systems, 2012.
- [9] Yogdhar Pandey, Darshike Lothe, "Evaluating the usability and security of a spelling based CAPTCHA system", International Journal of Computer Science and Information Technologies, vol. 5(3), 2014.
- [10] Elie Bursztein, Matthieu Martin, "Text-based CAPTCHA strengths and weakness", ACM Computer and Communication security 2011 (CSS'2011).
- [11] Brain M.powell, Gaurav Goswami, Mayank vasta, Richa singh, Afzel Noore, "FgCAPTCHA: Genetically optimized face image CAPTCHA," IEEE Access, volume 2, 2014.
- [12] Nitsha Payal, Nidhi Chaudhary, Param Nand Astya, "JigCAPTCHA: An advanced image based CAPTCHA integrated with Jigsaw piece Puzzle using AJAX", International Journal of soft computing and engineering, volume-2, issue-5, November 2012.
- [13] Jeff Yan and Ahmad Salah El Ahmad, "Usability of CAPTCHAs or usability issues in CAPTCHA design", Proceeding of the 4<sup>th</sup> Symposium on Usable Privacy & Security (SOUPS), 2008, ACM Press, pp. 44-52.
- [14] Ahmad El Ahmad and jeff Chan, Wai –Yin Ng, "CAPTCHA Design color, usability and security", IEEE computing society, 2012.
- [15] Carlos Javier Hernandez- Castro, Avturo Ribagorda, "Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study", ELSEVIER computer & security 29(2010) 141- 157.
- [16] The official CAPTCHA website: <http://www.captcha.net>.
- [17] Different types of visual CAPTCHA available at: <http://www.cs.sfu.ca/~mori/research/gimpy/>
- [18] Google CAPTCHA by best web soft at: <https://wordpress.org/plugins/google-captcha/>
- [19] Baljit singh saini, Anju Bala, "A review of bot protection using CAPTCHA for web security", IOSR journal of Computer Engineering.
- [20] Video base CAPTCHA available at: <http://www.cnet.com/news/stanford-university-researchers-break-nucaptcha-video-security/>
- [21] ESP PIX image available at: <http://blog.steren.fr/2011/02/20/re-captcha-nanojobs-and-gwap/>
- [22] Kitten Auth CAPTCHA available at: <https://thepcs.py.com/kittenauth/>
- [23] ASIRRA CAPTCHA available at: <http://research.microsoft.com/en-us/projects/asirra/>
- [24] Neha Chandrakant Mutha, Dr. Samidha D. Sharma, "3d Handwritten Animated Captcha Algorithm: Web Security", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 10, October – 2013, ISSN: 2278-0181.
- [25] High security CAPTCHAs standards: <http://solvemedia.com/security/>
- [26] Audio based CAPTCHA Image available at: <http://arstechnica.com/security/2008/12/computer-scientists-find-audio-captchas-easy-to-crack/>