Chaac -Captcha: An Improvisation of Graphical based Captcha with Dynamic Random Misrepresentation for Discrimination Between Human and Machine

Mr Mir Aman Sheheryar
M.Tech. (CSE) Student,
Department of Computer Science and Engineering,
School of Engineering and Technology, Sharda University,
Greater Noida, Uttar Pradesh, India

Dr Ashok Kumar Sahoo
Associate Professor,
Department of Computer Science and Engineering, School of Engineering and Technology, Sharda University,
Greater Noida, Uttar Pradesh, India

***Abstract*-- -In this paper, a CAPTCHA system is presented with dynamic misrepresentation which is provided by an offbeat, adequate and capable algorithm. Undoubtedly, it is believed that the ability of word text in CAPTCHA design depends on the recognition by humans. CAPTCHA which is the clipping of "Completely Automated Public Turing test to tell Computers and Human Apart" has been keeping on changing with rapid enhancement in the study and technology. In this paper, the concepts of etymological behavior of existing CAPTCHA systems are explored, and a new doctrine approach to build up the CAPTCHA is proposed. The proposed CAPTCHA designed in two variants (hard and easy) where hard variant of development withstand the intrusion. This results in a CAPTCHA which is stormy conditions, and is unrecognizable to current day optical character recognition up to 99% and at the same go is easy for human recognition. There by increasing the time to decode for automated scripts (bots) and remain much easy for human study when it comes to readability.**

Keywords-- CAPTCHA, Human Interactive Proof (HIP), Turing Test, Bot (Robot), Cryptography, Image Processing, Confidence Interval, Covariance.

I. INTRODUCTION

With the current growing trend of internet usage across the globe, one thing that pings the mind of every user is the security. Thus security is the primary concern when connected to the internet by any means. The security over internet is said to be best achieved when it envelops with Confidentiality, Integrity and Availability [1]. Hence the security concern over internet paved the way for CAPTCHA technology, that work in a procedural way to protect and provide tool against automated bots.

Verification forms the bedrock of secure system. Most of the activities performed on internet these days involve the exchange of users credential information so as to precede forward. Thereby going through any online activity, the user is subjected to verification in order to maintain his/her stand. In case if the vital information is provided to illegal authority, the whole protection of the system will break down. To achieve the versatile protection over internet first and the foremost thing comes to mind that the user interacting with computer to use internet is a Human and not an Automated script (Bot). In order to provide proper justification by user being Human or not, John Langford, Nicholas J. Hopper and Luis Von Ahan [2] introduced the term CAPTCHA at Carnegie Mellon University in the year 2000. The CAPTCHA has been made and implemented to withstand illicit actions from the automated bots. Up to now enormous CAPTCHAs have been developed that aid in security measures while interaction with the Internet.

The CAPTCHA involves Human Interactive Proof (**HIP** which is an acronym). **HIP** is a sequence of bounded instructions having the capability to point out difference between automated Scripts (Bots) and Human user.

The prime factor that governs the CAPTCHA technology is to judge between illegal and genuine users that is achieved through "Turing Test". In Turing Test, examination is being performed between two users. First user is Human and second is an automated script (bot). In return to give up the response, both users act to be same (Human) and similarly work upon to confuse the CAPTCHA system. Thus by judging and analyzing the response filed by the user, the CAPTCHA system has to distinguish between the two users which one is legal (Human) and the other is illegal (bot). Overall, the CAPTCHA is extensively used to defend the online assets from misuse by Automated scripts (bots). CAPTCHA is a summed up approach consisting an image processing and cryptography of simple and organized components that include characters and Images. CAPTCHA's are presented in image and the image is built by merging the characters (alphabets) and digits (numbers) in sequential or in random order over the meaningful or meaningless background. Using various methods, the purpose of CAPTCHA is to misrepresent and present to user for response, with misrepresentation the Optical Character Recognition (OCR) of an image and becomes difficult to obtain for present day OCR's [3].

Here a general approach for creating “Chaac-CAPTCHA’s” also pronounced as C-CAPTCHA is proposed. The C-CAPTCHA system is named after the Mayan god of storm/Rain that in classical Mayan civilization is known as ‘Chaac’. Chaac-CAPTCHA creation is based on random selection of characters and digits involving varying misrepresentation resulting in stormy conditions in the system. Hence the name Chaac is introduced in the proposed method. Also the user will be able to see through the C-CAPTCHA’s, while the computers (bots) will face tough time to see what they are. Here in section (I) the existing work in this field are introduced, section (II) provides theoretical foundation, and in section (III) the proposed algorithm for generation of Chaac-CAPTCHA is given. Lastly in section (IV) the outcome analysis followed and in section (V) conclusions drawn are discussed. And finally in section (VI) some light on future scope is presented.

II. EXISTING WORK

In order to devise a good CAPTCHA model it mandatory to explore the advantages and drawbacks of existing schemes on the basis of time complexity, strength, and technology used so as to acquire the desired outcome. As there have been lots of issues regarding automated scripts (bots) that is why security measures are taken into consideration while using web based applications giving call to CAPTCHA. Various techniques have evolved over the growing years to make CAPTCHA strong and versatile. The same is put forth in different research pursuers.

The author (Philippe Golle, 2008) [4] suggested the classifier which is efficient about 83 percent correct to tell accurately cats and dogs are apart that are used in *Asirra*. The proposed classifier is joint venture of support machine classier which is trained on color as well as texture features out of image. This classier allowed them to solve a 12-image Asirra challenge in automatic manner and the probability of error was 10:3%. Their study also looked into the attacks on partial credit and token bucket algorithms presented in “Asirra which is a CAPTCHA that exploits interest-aligned manual image categorization”. The contribution suggested the safe lookups of Asirra parameters for deployment.

The researchers (Datta et al, 2009) [5] in their research mentioned the mechanism to evade antagonist to perform network attacks that include Denial-of-Service which led to the scarcity of resources, they have stressed upon exploiting the limitations so as to protect automated network attacks potentially. Their work introduced the concept of recognition ability between humans and machines to distinguish images on the basis of their varying parameters with healthy as well as in noisy conditions .They study the application of controlled distortions on the basis of strength and varying nature and the effect on machine and human recognizability. It was observed that human recognizability depends upon the vast user study while as machine recognizability is based on storage in memory via content based image retrieval (CBIR) and some matching algorithms. For this, the devised technique is named as imagination which on testing and comparing between the different CAPTCHA classes showed the behaviors which are as fallowed :- *Automated Solutions*- Unlikely because of design, *Randomization* –High, *Input modality*- Mouse, *Time taken to solve*- Quick ,*Chance of Human failure*- Medium because of distortion attack trade off , *Educational Bias*- Low. Thus their proposed model has the advantage of posing Hard AI problems for image recognition by avoiding the shortfalls of the other image recognition systems. However, their good efforts to control the threat of getting solution to AI problem is extremely low .

The authors (Truong et al, 2010) [6] introduced the liable CAPTCHA by giving interactive CAPTCHA which is also called as i-CAPTCHA that is capable of defending the third party intervention. In their proposed approach, the user have to undergo through conventional testing procedure as in conventional CAPTCHA, but herein the time elapse is measured between server and client connection authentication that lays the basis to judge between the appropriate user which is legitimate. This idea is based on the text based CAPTCHA. For the implementation phase, they provide the user with image CAPTCHA to begin with. The CAPTCHA implementation was introduced while taking ICMA (Instant Messenger CAPTCHA Attack) architecture in consideration that contain two components attack script and IM connector where attack script is used for attacking particular website and IM enables instant messaging. Though good mechanism were proposed in their approach, yet it bears some limitations, as in case of an impaired user who suffers from slow reply time and protection against character recognition attack. However this approach provided a good basis to thwart the third party intrusion attack.

The researchers (Bursztein et al, 2010) [7] investigated over the need of CAPTCHA and efficiently solvable by humans in either of the variants whether it is in audio or in visual form. For their study, they took 31700 CAPTCHAs which were acquired from 20 popular CAPTCHA schemes among which 12 were image based and 8 are audio based. Their study, after analysis, reveals that CAPTCHA are difficult for humans as far as their intellectual capacities are concerned regarding different schemes. It has been seen that that audio CAPTCHA are difficult as compared with text and image for humans.

The authors in (Gao et al, 2010) [8] introduced the new approach against the text based CAPTCHA which they named as jigsaw puzzle. Jigsaw used the matrix concept of dividing the image into equal chunks , the image is equally distributed in the chunks and presented as the test for user to appear where the user have to arrange the jiggled image to formulate the actual image by swapping the matrix chunks with respect to their positions the swapping between the chunks. It is based on the edge matching technique. This approach proved good for human solving CAPTCHA with quick and accurate but the computers lack and rarely can. In their work, they presented the puzzle of two distorted chunks and by using image carefully rather than random selection of image for puzzle generation.

(Zhang, 2010) [9] introduced his own procedure of CAPTCHA and named it “Zang’s CAPTCHA”, using “Rich Internet Application” (RIA) with intelligent stimulus-response. The research was developed by using the scenario of flex and J2EE on client and server side respectively. After going through experimentation, it was seen that this formulated solution proved difficult for bot but easy to humans and resisted various attacks. The research done in this work also elaborated the architecture used in implementing RIA for intelligent stimulus response.

The authors (Lin et al, 2011) [10] introduced the drawing CAPTCHA in mobile scenario. The method is straight forward but it is not liable to depend upon. The research conducted by them have put forth the two main proposal an erosion-based CAPTCHA, breaking set of instructions (algorithm) that is successfully capable of attacking the drawing CAPTCHA for portable devices and second the new CAPTCHA called as zoo CAPTCHA. zoo CAPTCHA proved hectic for bots that are real concern inexistence.

The researchers (Te-En Wei et al, 2012) [11] gave the rejuvenating aspect of CAPTCHA. A new CAPTCHA scheme namely geoCAPTCHA. In this, the personalized content is used to prevent third party human attack. The personalized content used is taken as geographical information. This geoCAPTCHA has introduced the 3D viewing of images. Besides the new approach, their work also laid emphasis upon the various types of attacks that include “the third party attack, the botnet attack, vertical segmentation, Binarization , key-loggers, Man in the middle attack, phishing attack and Hidden camera attack. The geoCAPTCHA uses the location such as street preview, landmarks, scene images which are identified by user at selection time and are pre-registered by the server at the time of registration. For the authentication process, the user needs to rotate the challenge view to match the pre-registered image that is stored in the authentication server database. The calculation and comparison of street view challenge and pre-registered street view image is done by taking out their hash values and if the hash value matches the login is allowed and success is achieved. Their proposed approach provided good improvising over Google image CAPTCHA and is assumed to be suitably used in cloud authentication.

The authors (Mohammed Korayem et al, 2013) [12] gave the learning feature visualization on image based Avatar. In their research, they suggested the accuracy improvisation in Avatar algorithm using computer vision. On their findings, they achieved the features of computer vision which yields accuracy that are competitive and supersede human performance. In their procedure, they make the use of publicly-available “ICMLA Face Recognition Challenge dataset”, and applied straight forward visual recognition. They summed up their work after getting appropriated results in their experimentation and stated that the bots were unable to crack the proposed method in Avatar. However, if modern machine learning techniques are implicated there are possibilities that automatic visionary machine can break through.

(Kouritzin et al, 2013) [13] in their research proposed the efficient and effective algorithm for generating CAPTCHA using random field simulation. They introduced a general method for generating their proposed CAPTCHA and named it as “KNW CAPTCHA” which is pronounced as know-CAPTCHA by using random field simulation which supersedes the CAPTCHA used today. In their proposed work, they used an efficient algorithm to trigger a new KNW-CAPTCHA based on the parameters in Gibbs like manner, which proved that the proposed scheme developed a scenario of being good separator between computers and humans. Also facts were provided to check out the difficulty for computer programs to intrude through breaking and analysis to restrain segmentation and OCR attacks. They also discussed some target attacks and showed how to choose variants based on empirical results automatically. Further some methods were shown to Harden the proposed technique in case some loopholes are encountered. In their research, they used only black and white coloring while generating the samples. But, in future this technique could be extended to use different grey levels. However, the challenging thing would be adjustment of random sample generation and parameter estimation used to be extended up to that where from one can maintain and improve readability.

In (Obimbo, et al, 2013) [14] their paper intended to forward the solution to the growing problems related to CAPTCHA, so they introduced the CAPTCHA-All which is an improvisation to text based CAPTCHA. In their work, they presented the image scene of complex nature with combined task to user to solve, which involved the identifications of object in an image with simple click which proved good and less hectic for the user. On the other hand, it is more hectic for automated machines. Evaluation of the system had the great difficulty for attacker using brute force attack.

The researchers (Yamaguchi et al ,2014) [15] in their work figured out the limitations and vulnerabilities of the American developed CAPTCHA known as “we the people”, that is based on quiz session that prove hectic to solve for naïve users and was simple to attack. In their study, they introduced the verbal based CAPTCHA style which work upon to choose more significant option that is semantically and syntactically correct. Their idea came for the people bearing visual impairments so that they shall not suffer while appearing for questionnaires, This led to the concept of construction of new CAPTCHA against that of American based white house which lack the ability of naturalness among sentences.

The authors (Powell et al, 2014) [16] proposes fgCAPTCHA, that is a new image based CAPTCHA that uses face detection as the test. Their approach ropes the fact that humans are good when it comes at recognizing faces but at the same time this can be challenging for computers when distortion is applied. The approach is primarily developed for the touch-based input methods of mobile devices but is also compatible with point-and click techniques of traditional desktop and laptop computers. fgCAPTCHA is suitable for multilingual applications, unlike to address the usability shortcomings of existing implementations.

The authors (Shubhangi et al, 2015) [17] focused on new CAPTCHA generation methods in view of existing CAPTCHAs. They introduced the graphical CAPTCHA mounted on top of text versions but in mirrored and backtracked to confuse obfuscate users using the service over web. The introduced approach used in their work is based on certain architecture. Hence their proposed methodology consist an architecture having four modules which include user for login, server for authentication, and database for retrieving and graphical password for input password.

In a work by authors (Kaur and Behal, 2015) [18] proposed the new CAPTCHA model which is based on text in which they used varying parameters of text at one go that include change of font style, change of alignment, change of position and randomly generated words. On performing test their proposed system, they achieved the success rate of 95% when checked through OCR.

III. PROPOSED METHODOLOGY

A new method is introduced in this research for the Implementation of the Chaac-CAPTCHA system which is based on an improvisation of graphical based CAPTCHA systems. The proposed CAPTCHA system is unbreakable, more secure, and more robust as compared to existing CAPTCHAs. The improvised design of CAPTCHA is a combination of randomly generated selected symbols. The graphical based CAPTCHA includes two stages:

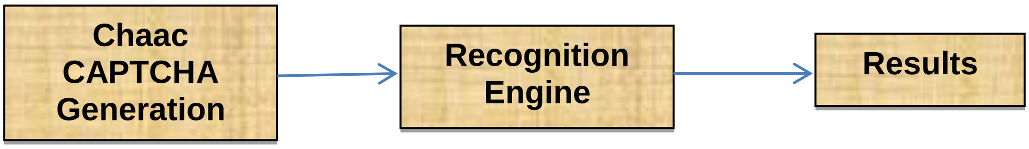

Figure 1: Showing the Block diagram of proposed scheme

a. *Chaac CAPTCHA generation*

First is CAPTCHA generation in which a random series of characters are generated. The generated series of characters is the combination of alphabets (upper or lower case) and numbers which prevent the dictionary attack in any system. Where in the presence of special characters are completely neglected so as to make the system more user friendly and less error prone. On the basis of covariance, the alphabets and number are merged into an image to obtain the CAPTCHA. Chaac-CAPTCHA is generated in two variants easy and hard. Where Chaac-CAPTCHA easy (E) is generated without any misrepresentation and denoted as Chaac-CAPTCHA 'E and Chaac-CAPTCHA hard (H) is generated using misrepresentation to make background unintelligible and is denoted as Chaac-CAPTCHA 'H.

b. *Recognition Engine*

In second stage, a recognition engine [19, 20] which is designed to aid as a mechanism to improvise the confidence interval, where confidence interval comprises a range of certain values (interval) that act as a good estimate while performing the hypothesis test with 95% confidence level and time to decode associated with the recognition engine. Time to decode is the procedural phenomenon time to solve the problem in terms of efficiency of decoding. Having lesser time to decode is the only success for bots. In order to improvise this weakness, a certain amount of efficiency is raised as compared to existing systems. In the recognition engine, the system includes various modern day features of optical character recognitions, segmentations in order to verify versatility of the proposed system.

c. *Results*

The outcome obtained from the first two stages is obtained in this phase that helps in further analysis of the proposed system.

A. *STEPS FOR THE PROPOSED MODEL*

The CAPTCHA generation and evaluation process consists of certain variations of a string formed by using an algorithm by selecting appropriate perimeters.

The steps listed below provide a guideline for creation of a basic Chaac-CAPTCHA image. For the purpose, a created CAPTCHA image is of desired dimension and able to hold the text string.

Now the generation of CAPTCHA begins with the generation of background misrepresentation which includes varying dynamic pattern so as to defeat vertical segmentation.

1. Select the random digits and alphabets: Generate six random characters from the designated character set and/or digit set. Generally, the upper A-Z and lower case English alphabets a-z are used as character set and 10 numerals, 0-9, are used as digit set. A
2. CAPTCHA test which is case insensitive and uses both alphabets and numbers thus has total combinations of 36 characters while as the one which is case sensitive, has a total combination of 62 characters. While obtaining the character set, the covariance of same character to be in next generated set to 2%.
3. A String to six characters: After forming the string of six characters, individual characters are rotated in the range of ±30^0. The rotation of characters is delimited within the range in Euclidian plane.
4. Generate an image containing the rotated character set: Embed the above generated image in it.

5. Join the rotated image generated randomly in pairs and introduce the dynamic (varying misrepresentation) resulting in stormy condition.
6. Select the random digits and alphabets: Generate six random characters from the designated character set and/or digit set. Generally, the upper A-Z and lower case English alphabets a-z are used as character set and 10 numerals, 0-9, are used as digit set. A
7. CAPTCHA test which is case insensitive and uses both alphabets and numbers thus has total combinations of 36 characters while as the one which is case sensitive, has a total combination of 62 characters. While obtaining the character set, the covariance of same character to be in next generated set to 2%.
8. A String to six characters: After forming the string of six characters, individual characters are rotated in the range of ±30^0. The rotation of characters is delimited within the range in Euclidian plane.
9. Generate an image containing the rotated character set: Embed the above generated image in it.
10. Join the rotated image generated randomly in pairs and introduce the dynamic (varying misrepresentation) resulting in stormy condition.

B. PROPOSED ARCHITECTURE

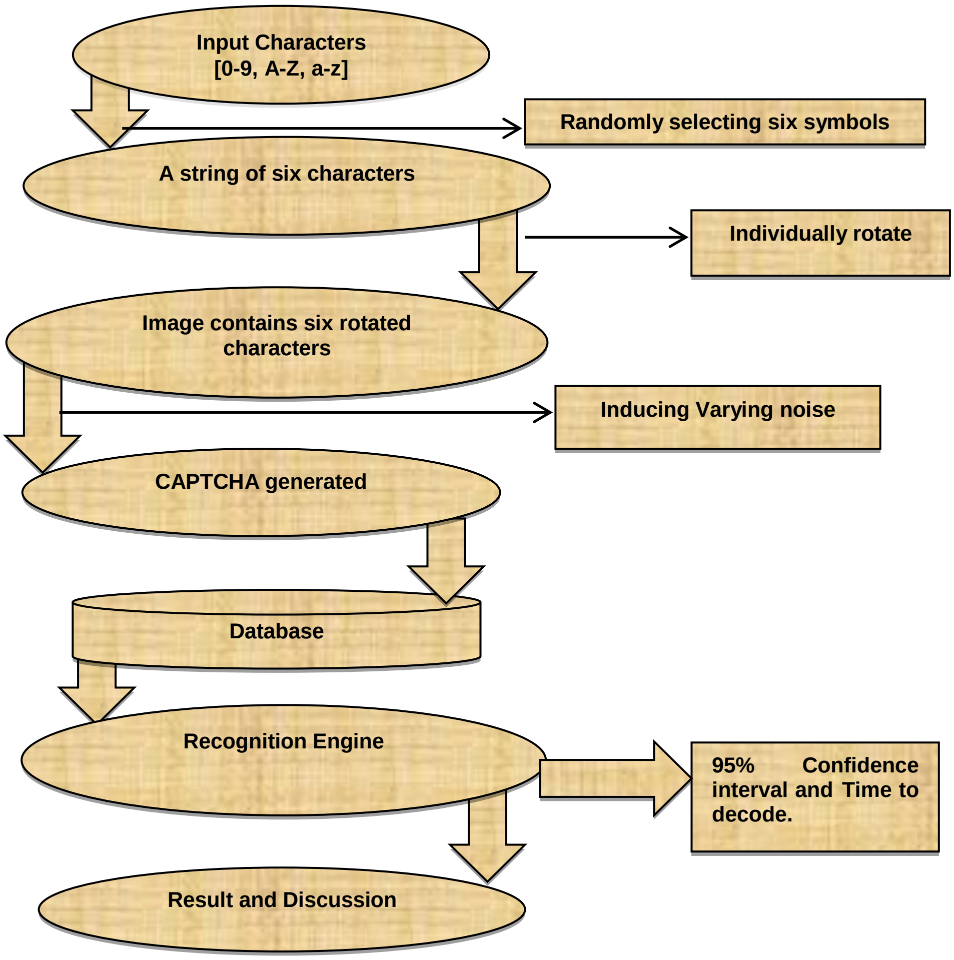
Input Characters
[0-9, A-Z, a-z]
Randomly selecting six symbols
A string of six characters
Individually rotate
Image contains six rotated characters
Inducing Varying noise
CAPTCHA generated
Database
Recognition Engine
95% Confidence interval and Time to decode.
Result and Discussion

Figure 2: Showing the Elaborated Diagram for Proposed Architecture

Here the misrepresentation obtained is by adding noise perimeters with varying pattern and of different type. The undetected background noise is generated using the following algorithm.

1. Choose the CAPTCHA image from the database.
2. Introduce the dynamic misrepresentation (noise) by choosing among the following parameters. For Chaac-CAPTCHA 'H formation.
3. Repeat the steps to 1- 4: Database of images are shown in figure 3 for Chaac-CAPTCHA.
4. One random image from database is given to the recognition engine. This CHAPTCHA is tested for performance analysis.

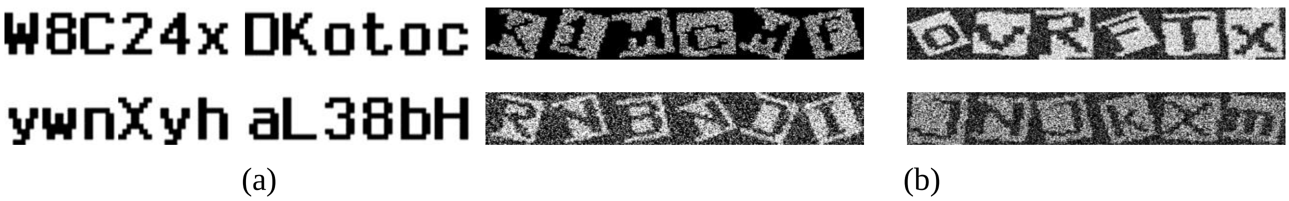

Figure3: Chaac-CAPTCHA examples. (a) Chaac-CAPTCHA 'E. (b) Chaac-CAPTCHA ’H, with dynamic misrepresentation.

Table1: The misrepresentation parameters

Speckle noise (in,'speckle',50)	Poisson, salt & pepper (in, 'Poisson') (in, 'salt & pepper',0.50)
Gaussian and Poisson (in,'gaussian',0.01,0.25) (in, 'Poisson')	Speckle ,Gaussian ,Poisson, salt & pepper (in,'speckle',10) (in,' Poisson') (in,'gaussian',0.01,0.15) (in ,'salt & pepper',0.10)

C. PERIMETER ESTIMATION

a) Change of noise:
In earlier CAPTCHA's static noise is used for every CAPTCHA. However, in the proposed Chaac-CAPTCHA, varying noise for each new CAPTCHA is introduced.

b) Change of Alignment: In the proposed CAPTCHA, the alignment of characters is varied. All alignments are not in a same line.

c) Change of position: Every time when a new Chaac-CAPTCHA is generated, the position of numbers and characters are varying.

d) Randomly generated words: No dictionary words are generated because they are easily breakable.

IV. RESULTS ANALYSIS

In this section the properties of Chaac-CAPTCHA and results produced are discussed. A system is proposed for Chaac-CAPTCHA 'E to cite the easy Chaac-CAPTCHA version (Figure 3a). This version is developed with no misrepresentation in background as well as among characters and is drafted to be easy while taking the basic bedrock attributes of Chaac-CAPTCHA in consideration. The other version is Chaac-CAPTCHA‘H and is developed with dynamic misrepresentation in background as well as dislocation among characters being used and is set up in routine (See Figure 3b).

A. *Chaac-CAPTCHA ‘E examinations*

In the easy version of the CAPTCHA design, the breaking rate of Chaac-CAPTCHA 'E is low, though it differs in case of modern OCR systems. Initially, test to check the resistance of our system is checked. In order to do so, the following examination is proposed. The examination is delimited among the occurrences of varying characters as per their presence in the new Chaac-CAPTCHA’ E image generated depending upon the period of occurrence of characters, so that one can relate how much the proposed system is readable to automated scripts (bots) and Humans. Thus to perform the procedure a recognition engine is designed. The Chaac-CAPTCHA ’E images are subjected through this recognition engine in order to achieve the desired outcome. The aim here is to achieve the reverse process of that by which Chaac-CAPTCHA 'E images are generated.

In order to do that the following procedure is followed.

1. Automated bot testing: involving OCR programs.
2. Filling the response questioner by human response.

In the first phase, the help from OCR programs available online are used (Tesseract, ABBYY reader) and other devised methods are used to crack down the CAPTCHA image. The procedure followed is in a way so as to give tough time to bot’s and also analyzed by the human approach. Keeping in view the need CAPTCHA ’E images are devised in a way so that it is easily distinguishable. (keeping in view its fundamental development)

1) Automated bot testing:

It involves OCR programs to read exact CAPTCHA image character and segmented the characters separately.

The CHAPTCHA image is presented to online OCR’s and the OCR systems performed the work of decoding string from the input CHAPTCHA. The set of 20 Chaac-CAPTCHA ’E are provided to perform this task. Please refer table 2 for outcome

'Tesseract' reader is unable to reorganize the characters properly and always produce wrong result which differs from the actual data in the images. 'ABBYY' reader also failed to load the image and displayed only blank outcome. The OCR program of the system only detected 70 characters out of 120 characters which sum up to build 20 Chaac-CAPTCHA ’E images of fixed length of 6. The OCR program obtained a success rate of 58.33% of decoding the characters among which only two Chaac-CAPTCHA ’E images are cracked fully. Please refer to table 2 for details.

Table2: Recognition engine outcome to Chaac-CAPTCHA' H

CAPTCHA Set	'Tesseract' Results	'ABBYY' Results	Self Recognition Engine Results
8	1 wrong character is detected	Nil	2 images were decoded
12	3 decoded but did not match with CAPTCHA image	Nil	3 characters from each image were detected
16	4 decoded but did not match with CAPTCHA image.	Nil	4 characters detected from each Captcha
20	6 decoded but did not match with CAPTCHA image	Nil	4 characters detected from each Captcha
Over all Observation	Failed to crack	Failed to crack	58.33% cracked

Feature extraction outcome:

To gain proper fragments of characters in CAPTCHA image, OCR's need to extract the characters [21, 22] individually so as to reorganize them. For details see table 3.

Table 3: Recognition engine outcome to Chaac-CAPTCHA' H

Action	Outcome
Maximally stable external regions (MSER) region	Detected
Canny edge and intersection of canny edge with MSRN region	Done
Edge grown along gradient direction	Least
Original MSER regions and segmented MSER regions	Partial
Text candidate before and after region filtering	Half
Visualization of text candidate stroke width	Partial
Text candidate before and after stroke width filtering	Half
Image region under mask created by joining individual characters	Done
Text region	Done

2) Filling the response Questioner by human response.

A set of 20 Chaac-CAPTCHA 'E images are given to 30 volunteers participating in this activity among which 15 are male and literate and15 are females also literate with different age groups. The volunteers took part in this study are anonymous and they are requested to sit in a room according to their willingness and no stipend was given. The study started by providing the volunteers the set of 20 Chaac-CAPTCHA 'E images each provided with different image set, that is, none of CHAPTCHA are repeated again and are instructed to solve them according to guidelines explained. The start and end time are noted down before and after filling the response form. On the basis of response filled by these volunteers the response obtained is summarized below in table 4.

Table 4: Human response to Chaac-CAPTCHA'E

No of image set	Start time	End time	No of correct responses	Time taken to decode single CAPTCHA	Time to reorganize single CAPTCHA
8	8:11:00	8:12:00	8	1 min	0.125 sec
12	8:12:00	8:12:37	10	1min 37 sec	0.114 sec
16	8:12:37	8:13:23	16	2 min	0.125 sec
20	8:13:23	8:15:23	16	2min 50 sec	0.125 sec

B. *Chaac-CAPTCHA 'H examinations*

In the proposed Chaac-CAPTCHA context, the probability of success is high. The proposed Chaac-CAPTCHA' H is presented with dynamic misrepresentation (varying noise). The obtained CAPTCHA is to be presented before literate probably English knowing (reading ability) Human having normal eyesight; if impaired then usage of proper correlating medical spectacles required. Thus the task is to calculate the probability of success by automated scripts (bots) and time to finish the test analytically. Keeping the base facts of Chaac-CAPTCHA 'E in foundation, the experimental results on the Chaac-CAPTCHA 'H is obtained with balance outbreak protection with readability.

The following experiment uses 1000 Chaac-CAPTCHA 'H images with varying misrepresentation.

The examination is delimited among the occurrence of dynamic misrepresentation as per their presence in Chaac-CAPTCHA 'H image generated resulting in stormy conditions. Also depending upon the period of occurrence of characters to figure out how much the proposed system is readable to automated scripts (bots) and humans. The process followed is described below.

1. Filling the response Questioner by human response
2. Recognition engine: Devised OCR.
3. Automated bot testing: involving Tesseract and ABBYY Fine reader.

1) Filling the response Questioner by human response

For this task a set of 20 Chaac-CAPTCHA 'H images are given to 30 volunteers participated in this activity among which 15 are male (literate) and15 are females (literate) with differing age groups. The volunteers took part in this response are anonymous and are requested to sit in a room according to their willingness and no stipend was given. The study started by providing the volunteers the set of 20 Chaac-CAPTCHA ‘H images with different image set among the 20 images displayed in front of them involved dynamic misrepresentation resulting in stormy condition so as to judge the ability of human user to distinguish the characters correctly and notify if the difficulty is faced by the user or not. The misrepresentation is employed in new hardened CAPTCHA is grouped under four subsets, for example, presence of misrepresentation with different forms, type and field values (presence level, threshold, density) the volunteers are properly instructed to solve them according to guidelines explained prior to the experiments, start and end time are noted before and after filling the response. On the basis of response filled by the volunteers it is seen that none of them failed to recognize the characters. The results obtained are listed in table 5.

Table 5: Human response to Chaac-CAPTCHA'H

No of image set	Misrepresentation type		No of correct responses	Time to reorganize single Captcha
5	Type1	Speckle noise	5	0.2 sec
5	Type2	Gaussian and Poisson	5	0.21 sec
5	Type3	Poisson, salt & pepper	5	0.21 sec
5	Type4	Speckle ,Gaussian, Poisson, salt & pepper	5	0.23 sec

From the above table it is clear that the outcome of hardened Chaac-CAPTCHA ’H is readable to human and did not halt the user to think what is displayed before him/her. The human response is correct if it matches the scrambled word otherwise it is incorrect if it differs. The average computation time obtained is around 0.2125 seconds for solving any Chaac-CAPTCHA ’H provided to the human user keeping confidence interval 95%.

2) Recognition engine: Devised OCR attack on dynamically misrepresented image

The recognition engine involved OCR intrusion attack that is decoding the scrambled presented image in simple form. The Chaac-CAPTCHA’H images are subjected through devised recognition engine to reverse the stormy conditions in order to achieve the desired outcome. The aim here is to achieve the reverse process by which the Chaac-CAPTCHA 'H images could be decoded.

Firstly CAPTCHA images are checked that the characters can be separated by OCR, that is, segmented into parts. For this, the set of Chaac-CAPTCHA ’H images undergo through OCR that is capable of dividing image into parts (segments). A set of 20 Chaac-CAPTCHA ' H images is supplied to OCR and waited for its accurate response. The result of the experiments is depicted in table 6.

Table 6: Recognition engine outcome to Chaac-CAPTCHA' H

Number of image set	Type of misrepresentation	Full image segmentation	No of characters obtained
20	Type1	No segmentation occurred	No character obtained
20	Type2	Unequal segmentation	2 characters obtained in each
20	Type3	No segmentation	1 character obtained in each
20	Type4	No segmentation	No character obtained

When an image CAPTCHA contains type1 misrepresentation no segmentation occurred and no character was obtained. When an image CHAPTCHA contains type2 misrepresentation with unequal division, it is observed on an average 2 characters are obtained. When an image CHAPTCHA contains type3 misrepresentation, no division occurred and only one character is obtained. Finally when an image CHAPTCHA contains type4 misrepresentation, no division is occurred and no character is obtained. For the omplexity involved in the generated Chaac-CHAPTCHA, please refer to figure 4.

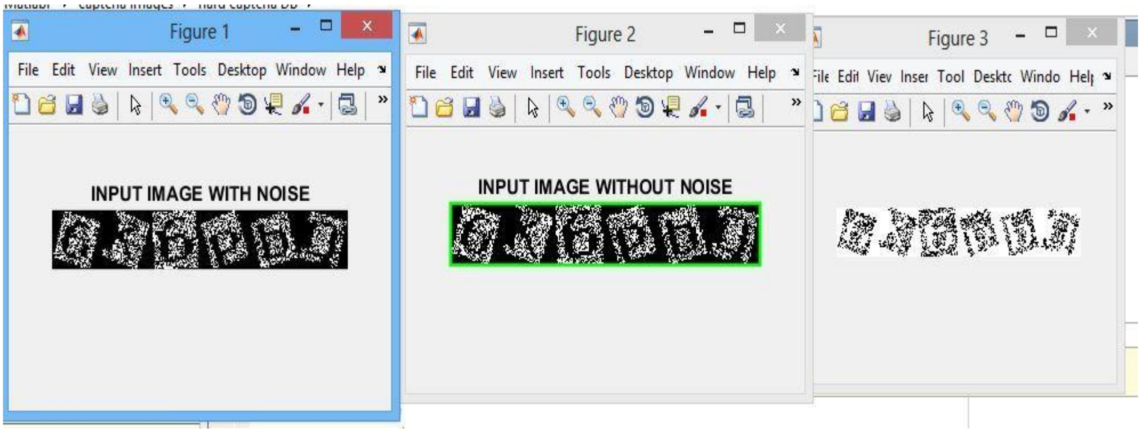
Figure 1
File Edit View Insert Tools Desktop Window Help
INPUT IMAGE WITH NOISE
Figure 2
File Edit View Insert Tools Desktop Window Help
INPUT IMAGE WITHOUT NOISE
Figure 3
File Edit View Insert Tools Desktop Window Help

TYPE1

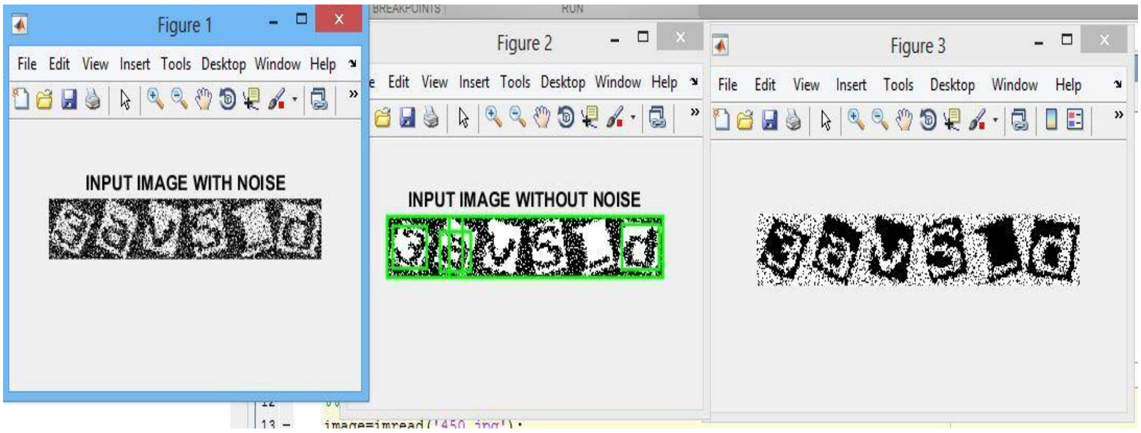
Figure 1
File Edit View Insert Tools Desktop Window Help
INPUT IMAGE WITH NOISE
Figure 2
Edit View Insert Tools Desktop Window Help
INPUT IMAGE WITHOUT NOISE
Figure 3
File Edit View Insert Tools Desktop Window Help

TYPE2

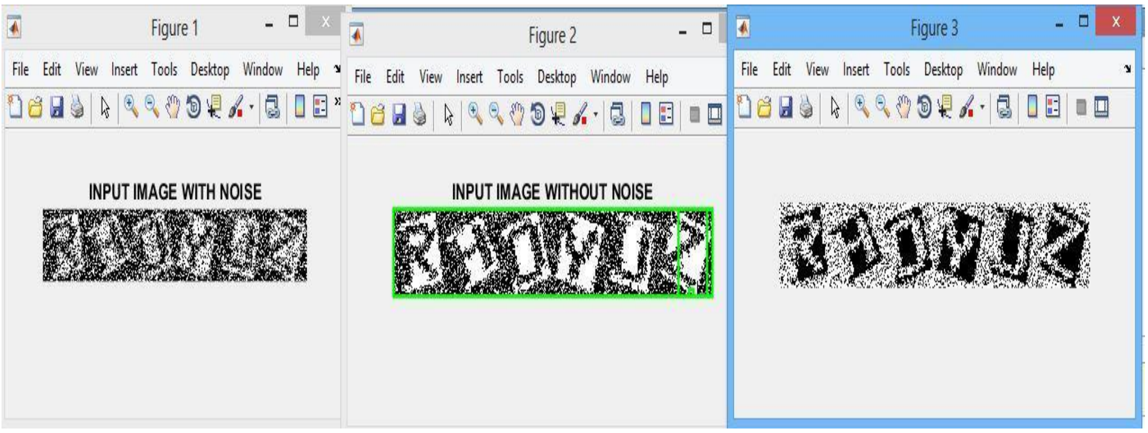
Figure 1
File Edit View Insert Tools Desktop Window Help
INPUT IMAGE WITH NOISE
Figure 2
File Edit View Insert Tools Desktop Window Help
INPUT IMAGE WITHOUT NOISE
Figure 3
File Edit View Insert Tools Desktop Window Help

TYPE3

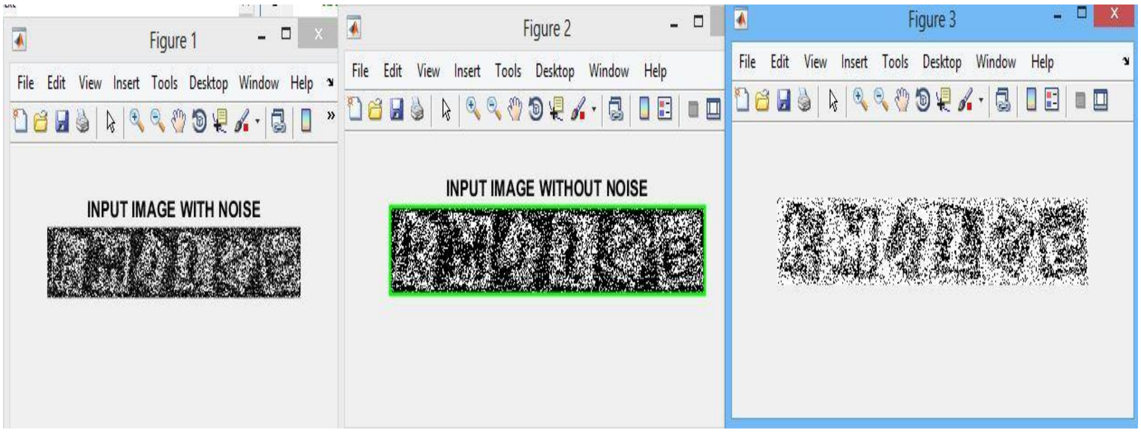
Figure 1
File Edit View Insert Tools Desktop Window Help
INPUT IMAGE WITH NOISE
Figure 2
File Edit View Insert Tools Desktop Window Help
INPUT IMAGE WITHOUT NOISE
Figure 3
File Edit View Insert Tools Desktop Window Help

TYPE4

Figure 4: Examples of type (1, 2, 3, 4) Chaac-CAPTCHA' H. Screenshots showing the recognition process by OCR when image is passed through recognition engine devised.

Feature extraction outcome:
To gain proper fragments of characters in CAPTCHA image, OCR's need to determine the characters individually so as to reorganize them. The detailed information is described in table 7 below.

Table 7: Recognition engine Feature extraction outcome to Chaac-CAPTCHA' H

Action	Outcome
Maximally stable extremal regions (MSER) region	Partially detected
Canny edge and intersection of canny edge with MSRN region	Partially done
Edge grown along gradient direction	Nothing detected
Original MSER regions and segmented MSER regions	Nothing detected
Text candidate before and after region filtering	Nothing detected
Visualization of text candidate stroke width	Nothing detected
Text candidate before and after stroke width filtering	Nothing detected
Image region under mask created by joining individual characters	Nothing detected
Text region	Nothing detected

3) Automated bot testing: involving Tesseract and ABBYY Fine reader.

Next, the Chaac-CAPTCHA 'H is compelled for discrimination between human user and automated scripts (bots) by bringing an “apples to apples” analogy of automated script (OCR) performance against human performance. To attain these outcomes, the Chaac-CAPTCHA 'H is subjected to undergo through 'Tesseract', 'ABBYY' Fine Reader for which the human responses are calculated and OCR responses have been devised and resolved certainty as before. The outcome is in table 8.

Table 8: Chaac-CAPTCHA' H HUMAN AND COMPUTER PERFORMANCE

Responses to Chaac-CAPTCHA' H	No of image set provided	95% Confidence Interval	Time to solve complete set	Time to solve one CAPTCHA
HUMAN RESPONSE TO C-CAPTCHA' H	1000	0.970±0.033	125 minutes	0.125s
Recognition Engine (Devised OCR)	1000	0.010±0.002	failed	N/A
Tesseract	1000	0.000±0.000	Failed	N/A
ABBYY	1000	0.000±0.000	Failed	N/A

By analyzing the outcome mentioned in the table7 it is clear that neither OCR modern program is able to crack through (decode) in order to perceive any of the Chaac-CAPTCHA' Hs. While as human user do remarkably well to perceive them. The modern OCR programs sporadically perceived any of the characters present in the image string (word).

Thus combining these outcomes obtained and taking them together with the outcome obtained in section IV -A, administer strong confirmation that the Chaac-CAPTCHA 'H evades the automated attacks confidently while maintaining simplicity for human user by being solvable quickly and easily. Notably it is determined that the times taken by humans to figure out our CAPTCHAs are not backbreaking.

Comparison

The correlation between Chaac-CAPTCHA' H to the other popular CAPTCHA systems provides an audit the susceptibility (loopholes) of the proposed system. The outcome obtain are absolutely outstanding. In order to obtain outcome a total of 20 Chaac-CAPTCHA' H images are selected from the database and the procedure repeatedly performed in section IV against the reorganization engine and modern OCR's. It was observed that the proposed system withstand against cracking. Also it was observed that few characters were extracted from Type2 misrepresentation category. However, the overall decoding was not possible. From this fact the success rate achieved from the proposed Chaac-CAPTCHA' H system is 99%. Correlations of Chaac-CAPTCHA' H with previous CAPTCHA systems are summed up in table 9 and figure 5.

Table 9: Chaac-CAPTCHA' H Correlation with Outcome Of Different CAPTCHA Against Computer Performance

Sr. No.	Name of CAPTCHA System	Percentage of breaking
1	Gimpy	80
2	EZ-gimpy	91%
3	Secure image	100%
4	Cryptography	100%
5	Mega upload	63%
6	Chaac-CAPTCHA' H	1%

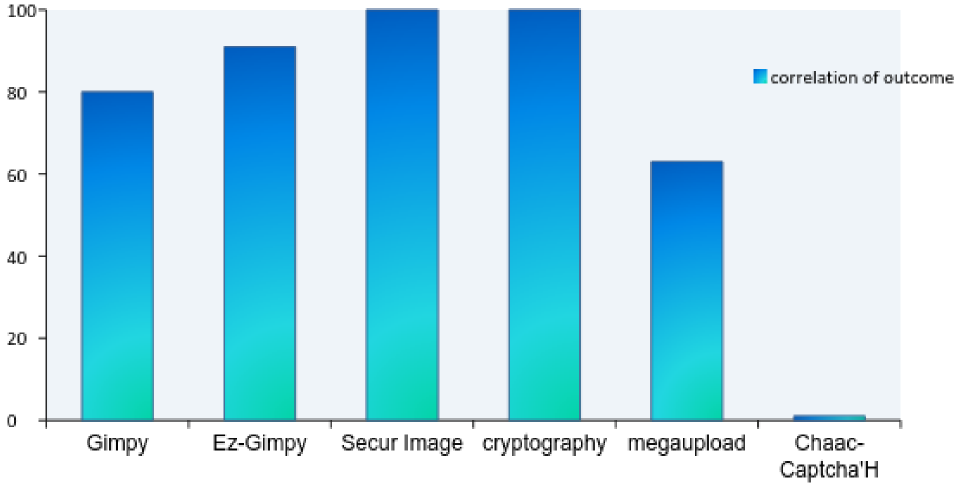

Figure 5: showing correlation of different CAPTCHA against Chaac-CAPTCHA' H

V. CONCLUSION

A new CHAPTCHA system is developed and high accuracy is achieved using a new approach of generating CAPTCHA's with dynamic misrepresentation, called Chaac-CAPTCHA's. That surpasses existing and contemporary CAPTCHA's in use.

Firstly a brief generalization for the need of security while being connected to network through internet, where CAPTCHA emerged to be most authoritative to discriminate between human and automated scripts involving appropriate accomplishing mechanism is discussed.

Secondly, the work done in preceding years so as to acquire an idea regarding the trends used in CAPTCHA scenario mechanism is summarized. For this, a well organized study survey after going through number of work's conducted by different researchers in the field of CAPTCHA's are presented. While going through preceding work it came to knowledge about the technologies, approaches, outcome of their conducted tasks. This is considered as research foundation in this area.

After that, a new approach of generating Chaac CAPTCHA which is "an improvisation of graphical based CAPTCHA with dynamic random misrepresentation for discrimination between human and machine" is developed. In order to determine the features of Chaac-CAPTCHA as best separator between human and computer (automated scripts), a number of tests are conducted.

Next, it is proved with sufficient evidence that Chaac-CAPTCHA is difficult for automated scripts while being easy for human users at same time by performing analysis of outcome. Two variants of Chaac-CAPTCHA are created. First one, without dynamic misrepresentation and named it as Chaac-CAPTCHA 'E another with dynamic misrepresentation and named that as Chaac-CAPTCHA 'H. In order to determine the strength of the proposed procedure, a number of experiments are conducted in the form of human readability, recognition engine (devised OCR) and also on online available modern OCR systems 'Tesseract' and 'ABBYY' Fine Reader and obtained the result. Thus it was proved that the proposed system is showing resistance to attacks (no decoding) and is much hard for automated scripts (failed segmentation) that at the same time proved easy for humans.

Finally, correlation between Chaac-CAPTCHA (Chaac-CAPTCHA 'H with dynamic misrepresentation) with the other CAPTCHA systems are performed in order to audit loopholes. The outcomes obtained are remarkable and outstanding. The overall decoding was not possible and the success rate of proposed CAPTCHA systems came to be 99% while comparing it with other CAPTCHA systems.

VI. FUTURE SCOPE

The system proposed can be well extended for the multilingual approaches. Also the display CAPTCHA image model bears the black and white colors that can be switched to color choices. Combined integration of audio and video can be looked upon to cater the need of human users from all outcomes. The main challenge will be to retain the easiness to reading while modification is introducing in dynamic misrepresentation in such a way that maintains or improve readability.

REFERENCES:

[1] Stallings, W. "Cryptography and Network Security", 4/e, Pearson Education India, 2006.

[2] Von A., L., Manuel Blum and John Langford. "Telling humans and computers apart automatically", Communications of the ACM, vol. 47, no. 2, pp: 56-60, 2004.

[3] Yan, J. and Ahmad Salah El Ahmad. "A Low-cost Attack on a Microsoft CAPTCHA", Proceedings of the 15th ACM Conference on Computer and Communications Security, ACM, 2008.

[4] Golle, P.. "Machine learning attacks against the Asirra CAPTCHA", Proceedings of the 15th ACM conference on Computer and Communications Security, ACM, 2008.

[5] Datta, R., Jia Li and James Z. Wang. "Exploiting the Human–Machine Gap in Image Recognition for Designing CAPTCHAs", IEEE Transactions on Information Forensics and Security , vol. 4, no. 3, pp:504-518, 2009.

[6] Truong, H. D., Christopher F. Turner and Cliff C. Zou. "iCAPTCHA: the next generation of CAPTCHA designed to defend against 3rd party human attacks", IEEE International Conference on Communications, 2011.

[7] Bursztein, E., Steven Bethard, Celine Fabry, John C. Mitchell and Dan Jurafsky. "How good are humans at solving CAPTCHAs? A large scale evaluation", IEEE Symposium on Security and Privacy, pp: 399-413, 2010.

[8] Gao, H., Dan Yao, Honggang Liu, Xiyang Liu and Liming Wang. "A novel image based CAPTCHA using jigsaw puzzle", IEEE 13th International Conference on Computational Science and Engineering, pp. 351-356, 2010.

[9] Zhang, W.. "Zhang's CAPTCHA Architecture based on Intelligent Interaction via RIA", International Conference on Computer Engineering and Technology, vol. 6, pp. V6-57, 2010.

[10] Lin, R., Shih-Yu Huang, Graeme B. Bell and Yeuan-Kuen Lee. "A new CAPTCHA interface design for mobile devices.", Twelfth Australasian User Interface Conference, vol. 117, pp: 3-8. Australian Computer Society, 2011.

[11] Wei, Te-En, Albert B. Jeng and Hahn-Ming Lee. "GeoCAPTCHA—A novel personalized CAPTCHA using geographic concept to defend against 3rd Party Human Attack.", IEEE Performance Computing and Communications Conference, pp: 392-399, 2012.

[12] Korayem, M. and David J. Crandall. "De-Anonymizing Users Across Heterogeneous Social Computing Platforms." The 7th International AAAI Conference On Weblogs And Social Media, 2013.

[13] Kouritzin, M. A., Fraser Newton and Biao Wu. "On random field completely automated public turing test to tell computers and humans apart generation", IEEE Transactions on Image Processing, vol. 22, no. 4, pp: 1656-1666, 2013.

[14] Obimbo, C., Andrew Halligan and Patrick De Freitas. "CaptchAll: An Improvement on the Modern Text-based CAPTCHA", Procedia Computer Science, vol. 20, pp: 496-501, 2011.

[15] Yamaguchi, M., Tatsuya Nakata, Hiromi Watanabe, Tatsuaki Okamoto and Hiroaki Kikuchi. "Vulnerability of the Conventional Accessible CAPTCHA used by the White House and an Alternative Approach for Visually Impaired People.", IEEE International Conference on Systems, Man and Cybernetics, pp. 3946-3951, 2014.

[16] Powell, B. M., Gaurav Goswami, Mayank Vatsa, Rajdeep Singh, and Afzel Noore., "fgCAPTCHA: Genetically Optimized Face Image CAPTCHA 5", IEEE Access 2, pp: 473-484, 2014.

[17] Hande, S. G. and M. S. Ali. "Enhancing the Security Using CAPTCHA as a Graphical Password", International Journal of Advance Research in Computer Science and Management Studies , vol. 3, no. 4, pp: 346-352, 2015.

[18] Kaur, K. and Sunny Behal., "Designing a Secure Text-based CAPTCHA", Procedia Computer Science, vol. 57pp: 122-125, 2015.

[19] Sahoo, A. K., G. S. Mishra and K. K. Ravulakollu. "Sign Language Recognition: State of the Art." ARPN Journal of Engineering and Applied Sciences. vol. 9, no. 2, pp: 116-134, 2014

[20] Sahoo, A. K. and K. K. Ravulakollu. "Vision Based Indian Sign Language Character Recognition." Journal of Theoretical and Applied Information Technology, vol. 63, no. 3, pp. 770-780, 2014.

[21] Sahoo, A. K. and K. K. Ravulakollu. "Indian Sign Language Recognition Using Skin Color Detection." International Journal of Applied Engineering Research. vol. 9, no. 20, pp. 7347-7360, 2014.

[22] Sharma, M., Ranjna Pal and Ashok Kumar Sahoo. "Indian Sign Language Recognition Using Neural Networks and kNN Classifiers." ARPN Journal of Engineering and Applied Sciences. vol. 9, no. 8, pp. 1255-1259, 2014.