HUMAN INTERVENED CAPTCHA

Thesis

Submitted to the

G.B.PANT UNIVERSITY OF AGRICULTURE & TECHNOLOGY,
PANTNAGAR-263145, UTTARAKHAND, INDIA

By

Pradeep Giri
B.Tech. (Computer Science and Engineering)

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

Master of Technology
(Computer Engineering)

August, 2016

ACKNOWLEDGEMENT

First of all I bow my head before 'God' who inspired me to face challenges of uneven times. All my sincere gratitude goes to him for the help he has given to me and his unfailing mercies over my life.

The author expresses his deep sense of reverence and heartfelt gratitude to Dr. Rajeev Singh, Assistant Professor, Department of Computer Engineering, Chairman of Advisory Committee for his invaluable guidance, constant encouragement, abundant counsel and his critical and constructive suggestions throughout the investigation and in the preparation of manuscript. The author is extremely indebted to him and wishes to thank him from the bottom of the heart.

With profound sense of gratitude the author expresses his warmest thanks to the members of the Advisory Committee, Prof. S.D.Samantaray, Professor & Head, Department of Computer Engineering, Prof. B.K.Singh, Associate Professor, Department of Computer Engineering and Prof. Ashok Kumar, Assistant Professor, Department of Information and Technology for their inspiring and constructive suggestions at every stage of this study.

The author tenders his sincere thanks to Dr. N.S. Murthy, Dean, College of Post Graduate Studies, Dr. S.P Singh, Dean, Student Welfare, and Dr. H.C. Sharma, Dean, College of Technology for their keen interest in providing the necessary facilities.

My abstruse regards goes to Ministry of Human Resource and Development, New Delhi for awarding me GATE Fellowship during the course of my M. Tech. degree programme.

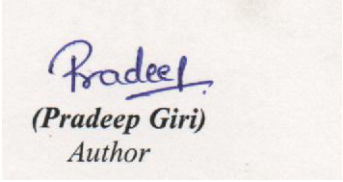
The author owes a very special word of thanks to his father Hoshiyar Giri, mother Shanti Giri and sisters Meenakshi Giri and Bhawana Giri for their boundless, generosity, everlasting inspiration, blessing abundant love and affection throughout.

Appreciations are also extended to my friends Ayushi Thapliyal, Ayushi Subhashini, Manmohan Singh Rautela, Sanjay Singh Bhandari, Shivam Mehta, and Neeraj Thakur for their encouragement and helping hands at various stage of the work.

I would like to give special thanks to Chanchal Agarwal for encouraging me and supporting me. She deserves my gratitude and if I forget to acknowledge her than it will be injustice with my soul.

This list is obviously incomplete but allow me submit that the omissions are inadvertent and I once again record my heartfelt gratitude to all those who helped me directly or indirectly in this endeavour.

Pantnagar
August, 2016

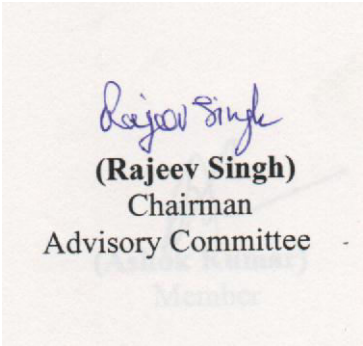


(Pradeep Giri)
Author

CERTIFICATE-I

This is to certify that the thesis entitled **"Human intervened CAPTCHA"** submitted in partial fulfilment of the requirements for the degree of **Master of Technology** in Computer Engineering with major in **Computer Engineering** of the College of Post-Graduate Studies, G. B. Pant University of Agriculture and Technology, Pantnagar, is a record of *bona fide* research carried out by **Mr. Pradeep Giri,** Id. No. **48141** under my supervision and no part of the thesis has been submitted for any other degree or diploma.

The assistance and help received during the course of this investigation and source of literature have been duly acknowledged.

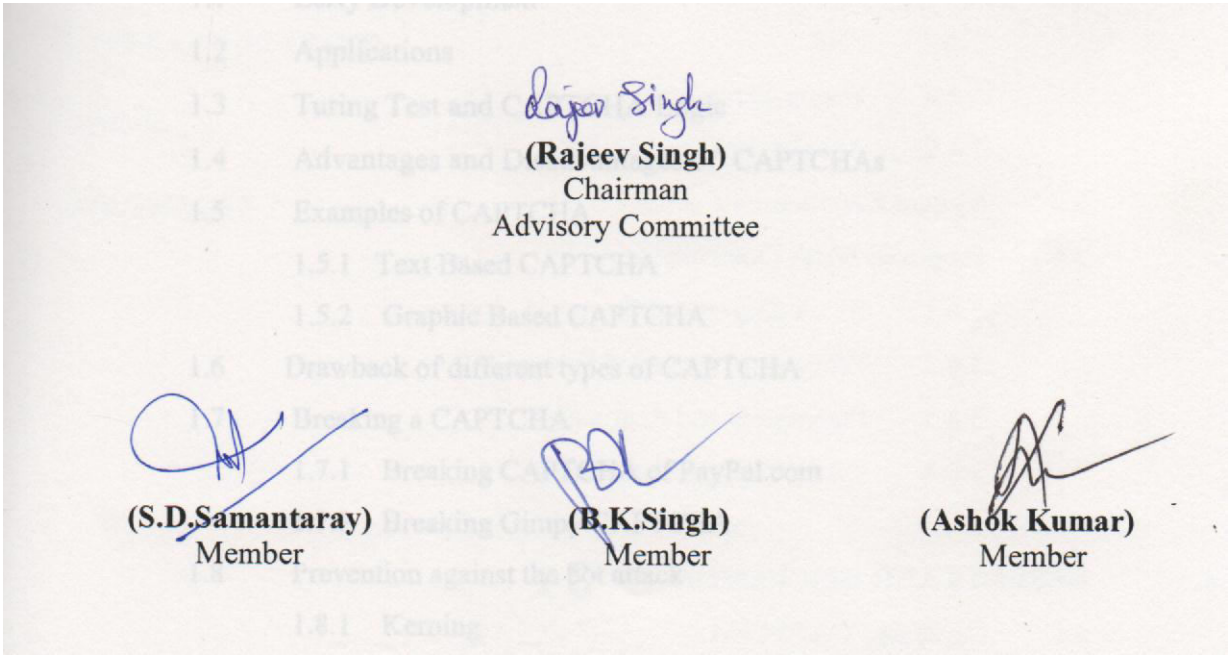


(Rajeev Singh)
Chairman
Advisory Committee

Pantnagar
August, 2016

CERTIFICATE-II

We, the undersigned, members of the Advisory Committee of **Mr. Pradeep Giri,** Id. No. **48141**, a candidate for the degree of Master of Technology in Computer Engineering with major in **Computer Engineering**, agree that the thesis entitled **"Human intervened CAPTCHA"** may be submitted in partial fulfilment of the requirements for the degree.



(Rajeev Singh)
Chairman
Advisory Committee

(S.D.Samantaray)
Member

(B.K.Singh)
Member

(Ashok Kumar)
Member

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CAPTCHA	Completely Automated Public Turing Test to tell Computers and Humans Apart.
CMU	Carnegie Mellon University
OCR	Optical Character Recognition
MSN	Microsoft Network
IEEE	Institute of Electrical and Electronics Engineers
URL	Uniform Resource Locator
AI	Artificial Intelligence
HIP	Human Interaction Proof
GHz	Giga Hertz
RAM	Random Access Memory
GB	Giga Byte
CSS	Cascading Style Sheet
HTML	Hyper Text Markup Language
PHP	Hypertext Preprocessor
WAMP	Windows Apache MySQL PHP
SMTP	Simple Mail Transfer Protocol
SQL	Structured Query Language
CPU	Central Processing Unit

Chapter 1 INTRODUCTION

CAPTCHA stands for Completely Automated Turing Test (**Ahn, L. V. *et.al.*, 2000**). It is encountered in one form or the other while using different accounts such as Gmail account, PayPal account or while commenting on any blog or while using our online banking account. CAPTCHA is basically a mechanism for identifying computerized bots over internet i.e. CAPTCHA identifies that the service is used by human or by any other automated program. Anyone using CAPTCHA embeds the automated program into the webpage at selective place and prompts the user to write the text generated by the automated program, thereby differentiating humans from bots.

CAPTCHA helps in many ways such as to stop the autonomous entries in website or to stop the computerized automated bots from doing unlawful activities on web page or stop spam attack in our password protected account.

GOOGLE is using its famous reCAPTCHA for blocking the automated spams, that automatically signup and degrade its services. EBay, Amazon and other e-commerce giants use CAPTCHA and protect themselves from flooding of un-necessary information over their website. Facebook and other social networking sites use CAPTCHA for limiting the fraudulent users. Banking sites uses it for authenticating the users.

CAPTCHA is also called 'Reverse Turing Tests' because they are programmed to identify the remote bots. CAPTCHA basically uses distorted tests i.e. the tests which can be recognized by the human but not by the computerized bots. As the time has passed new CAPTCHA techniques like Gimpy Captcha, MSN Captcha, Hotmail Captcha, Google's reCAPTCHA etc. have been evolved, which are very effective against automated bots.

1.1 Early Development

From early days of internet, user always wanted to make text reading and understanding difficult for computer or automated programs. The hackers usually break into the online journals or blogs and can post or change the sensitive data. So for securing the online data, they would use the simple technique. They replace the texts with the similar looking symbols and hence only human user would be able to identify them not the computerized bot. For example HELLO would be replaced as |-|3|_|_0. Though the

technique was very much elementary, but was able to fool the web crawlers, content filters and bots. This technique was eventually evolved as "133tspeak" (elite speak).

Spam bots can process thousands of web pages or web addresses and can send enormous amount of junk mails to the victims. Hence to stop such attack the web addresses should not be in bot recognizable state, addresses should only be recognized by human. Thus necessitating the need of protecting sensitive data online from automated bots.

1.2 Application

In recent years as the computing has evolved, the dependency of human beings over internet has also increased. There are many important areas where identification of bot or human is necessary. CAPTCHA has played an important role in web security, artificial intelligence, human computer interaction. This mechanism is able to differentiate between human or automated bot.

Many artificial intelligence problems are hard and unsolvable by the computers. For example computer is not able to understand the distorted texts, not able to understand images etc. These facts are utilized by the CAPTCHA. As the AI is evolving, Computers are becoming smarter and hence we are forced to use the smarter Captcha. These interesting facts were analysed by 'Ahn, L.V'. He concluded that it is win-win situation as either CAPTCHA is able to recognize human or bot, or if it fails means AI has solved one complex problem. Some of the important applications of CAPTCHA are:

a) Protection from fake account registration.
b) Protection from unnecessary flooding of useless data by automated bot.
c) Protection blogs or journals from fake comments by blocking bots.
d) Protection denial of service attack.
e) Protection from password cracking.
f) Protection from masquerading attacks.
g) Helps in preventing sensitive information of user and unnecessary posting of that information.

1.3 Turing Test and CAPTCHA Logic

In 1950 Alan Turing performed a test. In this test, user has to identify between human or machine. The test based on only 'text' related experiments and machine and

human are alike. If the user is not able to identify, which is human or which is machine, than the machine passes the Turing test.

Later on, a slight modification was done i.e. above roles were switched known as Reverse Turing test. CAPTCHA's are the slight modification of reverse Turing test. In CAPTCHA challenge is administered by machine and performed by human. The challenge is formulated such that it is not solved by machine but only by human. Hence it is able to differentiate between user and computerized bot.

Research team at Carnegie Mellon University was the first to give mathematical definition of CAPTCHA. They define CAPTCHA test V as (α, β) - human executable if at least α portion of human population has success greater than V.

CAPTCHA Logic is very simple:

a) Basic approach to generate CAPTCHA is, generate some random text, apply some random effects to it and convert it into image.
b) The machine generated CAPTCHA is provided to the user, and the user answers it.
c) The server checks, if the user entered answer matches or not. If the value entered by the user is empty or doesn't match, goto step 1 and regenerate the CAPTCHA.
d) If the user's entered answer is correct, continue with the application.

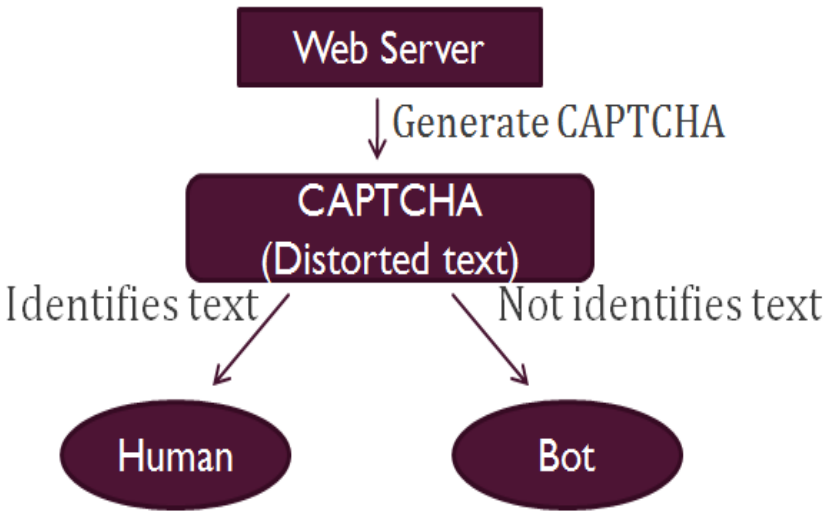


Figure 1.1: Reverse Turing Test

1.4 Advantages and Disadvantages

Table 1.1 Advantages and Disadvantages of CAPTCHA

Advantages	Disadvantages
• Helps in differentiating between bots and human.	• Distorted texts are sometime very much difficult to identify.
• Reduces spams and virus. • Helps from unnecessary flooding of information over internet.	• Not suitable for disable persons. • Time consuming
• Password protection; reduces risk of password decryption from brute force attack, as bots will be identified and blocked.	• Not always safe. As AI is improving day by day, the system may fail.
• Helps in making online polling system more genuine.	• Sometime browser compatibility issue may arise.
• Makes online shopping system safer.	

1.5 Example of CAPTCHA

1.5.1 Text Based CAPTCHA

Text based Captcha uses the letters, numbers and special symbols. The letters, numbers and special symbols form character set. Distortion is applied to the generated sequence of the character set. An image of the character set is formed, which is not recognizable by the bot. But humans are able to understand it properly. Some of the important Text based Captcha's are as follows:

The Gimpy CAPTCHA

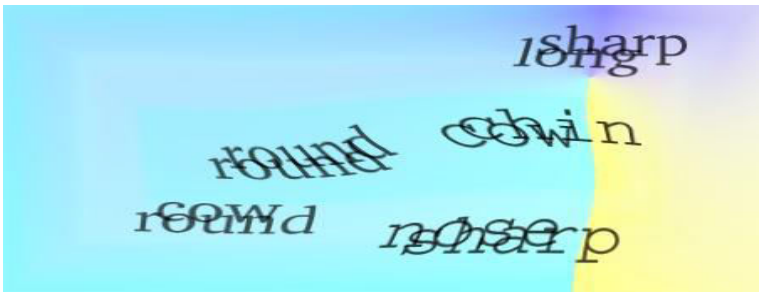


Figure 1.2: The Gimpy CAPTCHA (Greg M. et.al.)

The Gimpy(Ahn**, L. V.et.al.)** Captcha is one of the basic CAPTCHA. This CAPTCHA is based on the distortion level of the text. High distortion level makes it difficult for the user to understand it correctly. So a balance is made in deciding the distortion level. In Gimpy Captcha, seven different words are selected from the dictionary. The image is then created containing the distortion of these seven different words. The distorted Captcha image is then presented to the user and a direction or hint i.e. "type three words appearing in the image" is given to the user. The distortion level is applied in such a way that human is able to identify it but not the computerized bot.

Gimpy is very reliable Captcha. Most of the Captcha used in the web, directly or indirectly depends on it. Gimpy Captcha depends upon the difficulty level of Optical Character Recognition. Gimpy uses the fact that human is able to read the distorted text, but not the computerized bots. The Gimpy Captcha consists of seven or eight random words selected from dictionary. The distortion is applied to each of these words and asks the user to identify three or four words from them. The human user is capable of identifying the words correctly, whereas a computer program cannot. **(Sarika et.al.)**.

The Ez-Gimpy CAPTCHA

Figure 1.3: The Ez-Gimpy CAPTCHA (Greg M. et.al.)

The simplified version of Gimpy Captcha was developed by Henry Baird. He termed it Ez-Gimpy Captcha. It is used by Yahoo in Yahoo messenger in its sign up page. The Captcha uses a method, in which a single word is randomly chosen from the dictionary and distortion is applied to the word. The user has to identify the word. The method was not suitable and the Captcha was broken by the OCR technique (**Saini, B.S. et. al.**).

The MSN CAPTCHA

Figure 1.4: The MSN CAPTCHA (Saini, B.S. et. al.)

The Microsoft uses different types of Captcha for the services which are provided under MSN umbrella. These Captcha are termed as MSN Passport Captcha's. The MSN Captcha uses eight different letters i.e. upper case and numbers. The foreground colour of the Captcha is dark blue and the background colour is grey. Wrapping of the character is used to distort the text and produce the ripple effect. This distorted text is difficult for the computerized bots to identify. The MSN Captcha was not very much secure against the bot attack. The text image contains the distorted upper case letter and digits. High distortion level can also make it difficult for the user to identify. Therefore a balance is maintained in the distortion level, so that only human can identify the text not the bot. The Captcha was broken by Optical Character Recognition technique, as segmentation is applied to the image and hence bot can identify each individual character.

The HOTMAIL CAPTCHA

Figure 1.5: The Hotmail CAPTCHA (Saini, B.S. et. al.)

The Hotmail Captcha was not very much secure against the computerized bot attacks. The Hotmail Captcha consists of upper case letter, lower case letter and digits. Distortion level is applied to the character sets, so that only human can identify it not any bot. But this Captcha fails, as the Optical Character recognition (OCR) can segment the character set and this makes it vulnerable against the bot attack. The special thing about Hotmail CAPTCHA, than the Gimpy CAPTCHA is that it uses the combination of upper and lower case letters.

The Google's reCAPTCHA

Figure 1.6: The Google's reCAPTCHA (Ahn, L. V. et.al.)

Google's reCAPTCHA system is the advanced method, which uses an advanced risk analysis engine and adaptive CAPTCHA to stop the automated bots from entering into the system. It is the free service to protect the website from spam.

reCAPTCHA offers more than just a spam protection. The Google's reCAPTCHA is designed in such a way that, the human effort required to solve the CAPTCHA is utilized in some creative way. As the AI is evolving, CAPTCHA's are becoming smarter. Google's reCAPTCHA is one of them. Every time the CAPTCHA is solved the effort is utilized in digitalizing the books, annotate the image and building the machine learning data sets. Google's reCAPTCHA helps in making knowledge database of streets, which helps in making Google map more precise and complete. The image provided to the user is the one which is not easily understandable by the machine. Once the user enters the image text, it stores it in the knowledge database of the Google Maps. In the similar way, Google's reCAPTCHA helps in utilizing the human effort in digitalizing the books. The user is provided with the two texts, one is recognizable used to identify the user and the

next text is entered for machine learning. The words selected in the Google's reCAPTCHA are not read by the OCR. Thus step by step Google reCAPTCHA helps in book digitalization process. It can also be used for solving the hard AI problems, as Google's reCAPTCHA helps in building dataset for the AI problems. Some of its important features are as follows:

a) Google's reCAPTCHA works on the idea that the human effort required to solve the Captcha should be utilized in effective way.
b) Google's reCAPTCHA helps in making knowledge database of streets, which helps in making Google map more precise and accurate. The image provided to the user is the one which is not easily understandable by the machine. Once the user enters the image text, it stores it in the knowledge database of the Google Maps.
c) Google's reCAPTCHA helps in utilizing the human effort in digitalizing the books. The user is provided with the two texts, one is recognizable used to identify the user and the next text is entered for machine learning. The words selected in the Google's reCAPTCHA are not read by the OCR. Thus step by step Google reCAPTCHA helps in book digitalization process.
d) It can also be used for solving the hard AI problems, as Google's reCAPTCHA helps in building dataset for the AI problems.

For visually impaired persons, Google's reCAPTCHA also provide an audio CAPTCHA support. These audio Captcha are easy for the human to identify, while automated bots finds these audio Captcha's very much difficult. Hence bots are stopped. The audio Captcha are designed in such a way that these are very much hard for the bots to solve.

The Baffle Text CAPTCHA

Figure 1.7: The Baffle Text CAPTCHA (Ahn, L. V. et.al.)

The Baffle Text Captcha was designed by Henry Baird at University of California at Berkeley. This Captcha is the slight variation of the Gimpy Captcha. As the Gimpy Captcha uses the dictionary words, it uses any random letters to produce the pronounceable text. The text may have not meaning, it is just the collection of the letters. Then the distortion is applied to the text and passes to the user to guess it. This technique overcomes with the drawback of Gimpy Captcha and Ez-Gimpy Captcha. As the Gimpy and Ez-Gimpy Captcha use the words of the dictionary, it is a possibility that the clever bots can apply the brute force technique and identify the word and break the Captcha system. Therefore the designed Baffle Captcha uses random letters to form the word and then apply distortion to it.

1.5.2 Graphical CAPTCHA

Graphical CAPTCHA uses different images to identify the bots. This type of Captcha does not use text. It uses image or a combination of different images and a question associated with the image. The questions may be, identify the tree in the figure or identify the ice-cream, select the circles in the image etc.

The Bongo CAPTCHA

The Bongo Captcha program is unique type Captcha program that ask the user to solve the visual patter-recognition problem. The Bongo Captcha displays the series of two blocks of different patterns. The left series and the right series are differ in certain patterns. The user has to identify the unique pattern or find the characteristics that the two sets are apart.

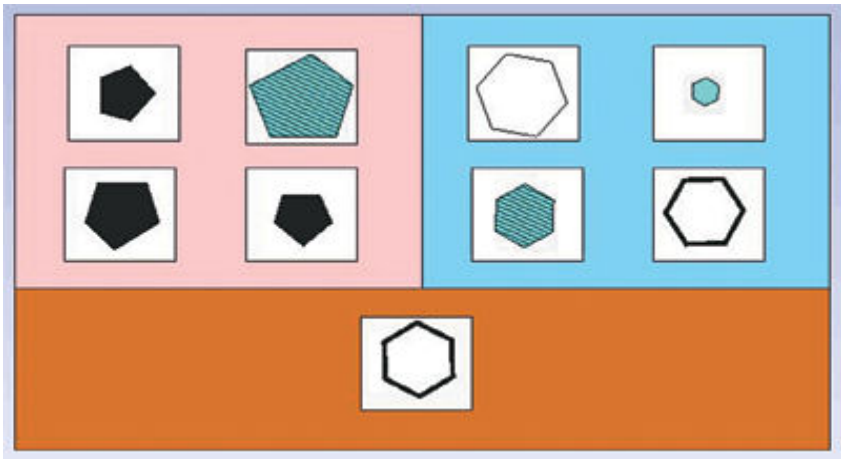


Figure 1.8: The Bongo CAPTCHA

The possible two series are given in figure 1.8. After seeing the two series blocks, the user is given with the different blocks and a question is asked that the given pattern belongs to the left series or the right series. The user passes the test if he answers it correctly and determines the patterns series.

The PIX CAPTCHA

Figure 1.9: The PIX CAPTCHA

The Pix Captcha program contains a large database of labeled images. All these images are pictures of the concrete objects i.e. cats, flowers, clock, car, water etc. The Pix Captcha program picks the name of the object randomly and based on the objects name, random images of the object are selected from the database. A collage of the image is formed and presented to the user. Based on this image a question is asked with the user. The questions may be identify the flower from the image or click the car or click the water etc. If the user answers, he passes the Captcha test.

However Pix is not a Captcha program. It is very easy to write the code for this type of program. All the code is freely available in the internet. Apply the brute force technique to search the image in the database and the label associated to the image. This problem of the PIX can be resolved. Apply distortion to the images in the database and makes it difficult for the automated program to find the distorted image from the database.

1.6 Drawback of different types of CAPTCHA

Table 1.2 Different types of Captcha and their drawbacks

Sl.No.	Different types of CAPTCHA	Drawbacks
1.	Text Based Captcha	1. User has problem in identifying the text image. 2. Multiple fonts and high distortion level, makes it difficult for identification of the character set. 3. It can be easily identified by the OCR technique, because segmentation of the image text is easily possible.
2.	Image Based Captcha	1. Users with vision problem are not able to identify the Captcha. 2. Sometimes the large images are not downloaded.
3.	Audio Captcha	1. Captcha is based on English vocabulary. Hence user should have firm command over it. 2. Characters or words having similar sound can cause problem.
4.	Video Based Captcha	A large video file of the Captcha may cause difficulty for the user while downloading it.
5.	Puzzle Based Captcha	A large amount of time is required to solve the Puzzle.

1.7 Breaking different types of CAPTCHA

As Machine Learning is evolving, the need for defining better CAPTCHA is also increasing. CAPTCHA's are being widely used by websites to stop the spam or bot attack. CAPTCHA are constantly under attack since 2000. Since then every effort is made to define the better Captcha. Breaking of Captcha is always a win-win situation. If a Captcha breaks, it means that hard AI problem is solved. Captcha's are always defined over AI problems, it means the defined CAPTCHA is a type of hard AI problem which cannot be solved by computer programs.

1.7.1 Breaking CAPTCHA of PayPal.com

The PayPal Captcha suffers from many weaknesses i.e. fixed font size, fixed font length, no distortion etc. Hence the Captcha is easy to segment. The algorithm developed to break the PayPal Captcha comprises of three steps. The image is pre-processed using simple cleaning techniques to remove the noise level from the image. The processed image is then segmented using vertical projection and split positions. Then four classification methods are implemented namely pixel counting, vertical projection, horizontal projection and template co-relation. The system was trained with the sample of 20 PayPal Captcha's to create thirty six training templates (one for each character: a-z, 0-9). The sample of 100 PayPal Captcha's were used for testing and about 88% success rate was achieved. **(Kluever, K. A.).** The steps are shown in the figure 1.10.

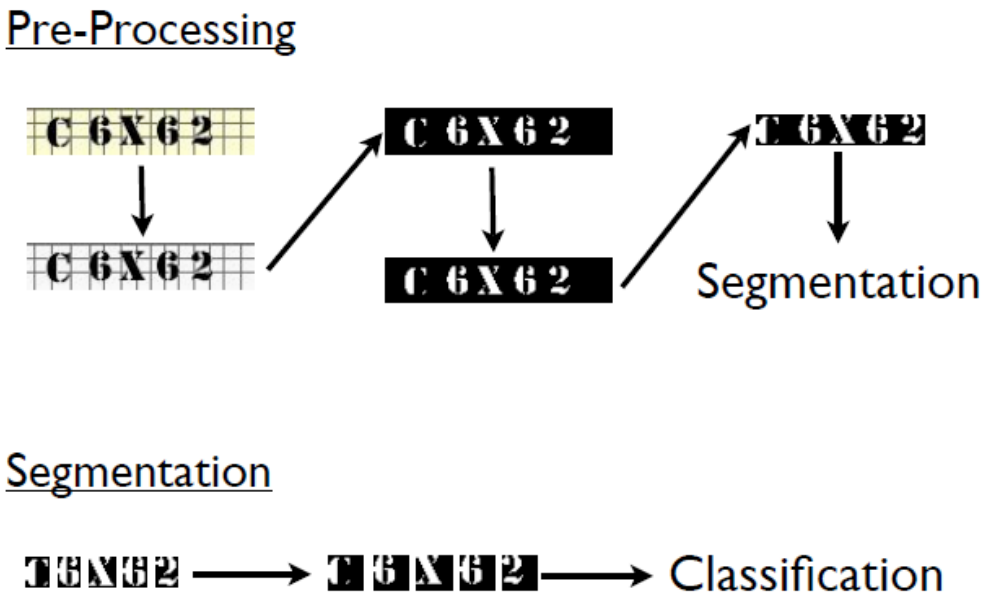


Figure 1.10: Breaking a PayPal CAPTCHA Process (Kluever, K. A.)

1.7.2 Breaking Gimpy CAPTCHA

The Gimpy Captcha is the basic Captcha from which all the Captcha's have been evolved. The Gimpy Captcha consists of seven or eight random words selected from dictionary. The distortion is applied to the word and asks the user to identify them. The breaking of Gimpy Captcha is four step processes. Firstly remove the distortion from the image by using simple cleaning mechanism. Then remove the background shade. Finally apply segmentation to the image and identify each word correctly.

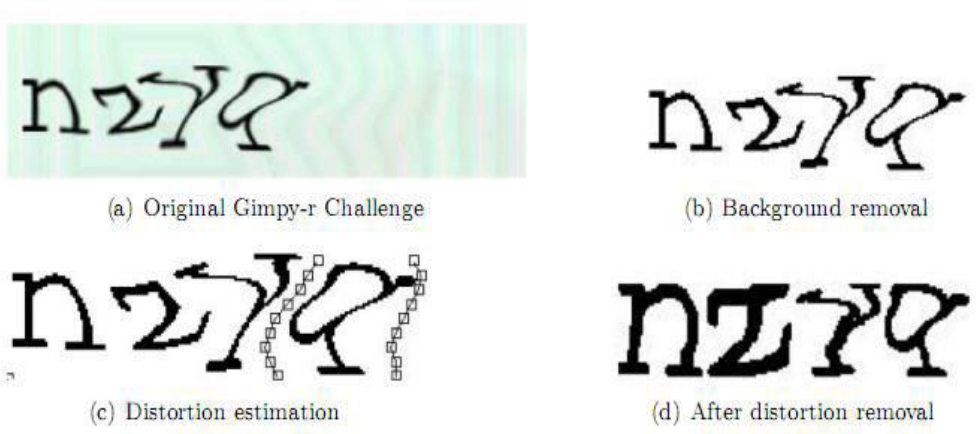


Figure 1.11: Steps in breaking the Gimpy challenges

1.8 Prevention against the Bot attack

1.8.1 Kerning

Kerning involves adjusting the amount of or removing the Space between characters. This increases the difficulty of isolating individual characters using OCR.

Figure 1.12: Kerning (Craig, M. S. et.al.)

1.8.2 Stretching

Figure 1.13: Stretching (Craig, M. S. et.al.)

In this method, random level of stretch or compression is applied to the object to increase the difficulty of OCR recognition.

1.8.3 Varying the font, size and style

Many types of OCR systems are tuned for the specific type of document that will be scanned and thus will have difficulty identifying text of varying fonts, styles and sizes.

Figure 1.14: Varying Font (Craig, M. S. et.al.)

1.8.4 Spatial Transformations

These types of transformations include Rotation of the object to the left or right or up or down and randomizing its location within an image.

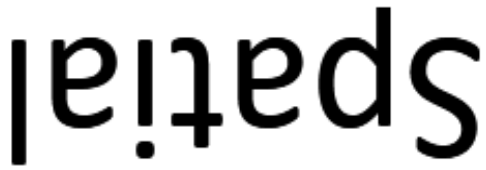


Figure 1.15: Spatial Transformation (Craig M. S. et.al.)

1.8.5 Lossy Image Compression

Lossy process makes it difficult for the automated bot to identify the text and hence secures against the bot attack.

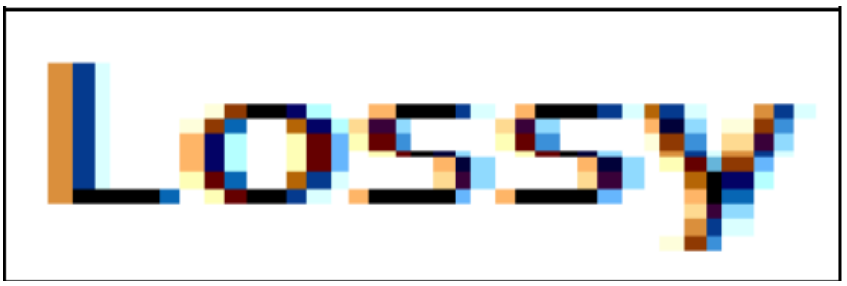


Figure 1.16: Lossy Image Compression (Craig M. Schow et.al.)

1.9 Motivation

Today almost every website uses Captcha for the bot identification purpose. Captcha's are used before sign in or sign up, so that only genuine user enters into the system, not any bot. CAPTCHA is used to stop the spam attack in the website. Captcha's have evolved a long way. As the AI is evolving, Captcha's are becoming smarter and more secure against the bot attack. Now Captcha's are used for multi-tasking feature also. For example, Google's reCAPTCHA not only identifies the bot but also help in the machine learning process.

The proposed system comprises of two CAPTCHAs. One is Automated CAPTCHA, which is used for the detection of online bots and second is Human Intervened CAPTCHA which is used to differentiate between genuine and invalid users. A genuine user is the one who is authorized to use the system.

The Automated CAPTCHA is based on randomly generated character sets. The length of the character set is also random generated. The user has to answer the question based on generated character set. The type of question is also randomly selected from the pool of questions. Too much randomness makes it safe against the bot attack, because bots are not able to identify the character set, its length or the type of question.

As Google's reCAPTCHA performs the multiple tasks i.e. identify the bot and helps in machine learning. In the same way, the thesis proposes a system of Human Intervened CAPTCHA. In this system multiple questions are generated by the admin and passed to the user. The answer of the question is not fixed; it changes according to the location, time and situation of the admin. The idea behind the Human Intervention Captcha system is that the human effort required to solve the Captcha is utilized in some creative way. The proposed HI-CAPTCHA not only differentiates between bot and user/human but also differentiate between genuine user and invalid user. Besides these, the proposed HI-CAPTCHA system can be used for e-polling, maintaining records, getting details of the user etc.

2.1 Literature Review

Literature review helps in understanding the research problem. It is consistent way of qualitative and quantitative study of the problem domain. Many IEEE documents and books of CAPTCHA have been referred for the problem domain.

Mark *et. al.* (1998) Martin Abadi, Krishna Bharat and Andrei Z. Border invented the CAPTCHA method. He developed a mechanism for the ALTA VISTA search engine, which identifies the bot and prevents from adding URL's to the search engine by the bot. They use a distorted English text which is presented to the user. If the user identifies the text, he is human otherwise bot. This system is even used today in web. Though the term CAPTCHA was not given by them, but the same mechanism is followed in CAPTCHA.

Ahn *et. al.* (2000) coined the term CAPTCHA. They explain the working and mechanism of CAPTCHA. The CAPTCHA security was explained along with its running demonstration.

Mori and Malik (2000) proposed the Gimpy and EZ-Gimpy technique for object recognition. This method is still famous and commonly used by Yahoo. They developed an efficient method of shape matching, which helps in identifying the word from the Gimpy and Ez-Gimpy image with a success rate of 92%.

Ahn *et. al.* (2003) claimed in their work that, they have invented the CAPTCHA. They proved that CAPTCHA is basically a hard AI problem and breaking of CAPTCHA always leads to the win-win situation i.e. either the CAPTCHA remains unsolved and helps in differentiating between bots and human or one of the hard AI problem get solved.

Ahn *et. al.* (2008) proposed that the human effort which is required to solve the CAPTCHA can be used for useful purposes. For example, it can be used to decipher the hard words which are not understand by the machine or by OCR and thus help in digitizing the books and which further helps in machine learning process. They demonstrated that their method can transcribe the text with 99% accuracy.

Kluever (2008) mentioned the breaking of PayPal system. In his work, he mentioned a process in which firstly an image is pre-processed and the noise is removed. Simple cleaning techniques are used for image pre-processing. This method basically deals with the removal of distortion level of the image. In this method the image is then segmented. Different classification methods are used such as pixel counting, vertical projection, horizontal projection and template co-relation.

The system was initially trained with the sample of 20 paypal.com Human Interaction Proof (HIP's) to create thirty six training templates (one for each character: a-z, 0-9). A sample of 100 paypal.com HIP's were then used for testing and about 88% success rate was achieved.

Yan and Ahmad (2008) analysed the security of the CAPTCHA designed by Microsoft. These CAPTCHA are used at Hotmail, MSN and Windows live. The CAPTCHA schemas were designed such that they become segmentation resistant. However one simple attack achieved the segmentation problem higher than 80%. It took 80ms to completely segment the problem with 1.86 GHz Intel Core 2 CPU and 2 GB RAM. As a result it was estimated that MS CAPTCHA scheme could be broken with the overall success rate of more than 60%. On the contrary, its design was so much robust that automatic scripts should not be more successful than 1 in 10000 attempts (success rate of 0.01%). For the first time it was concluded, that the CAPTCHA that is segmented resistant could be vulnerable to the simple attacks. The result showed that it is not a trivial task to design a CAPTCHA scheme that is both robust and vulnerable to the segmentation.

Gupta (2009) firstly proposed the use of numbers in the text based CAPTCHA. He mentioned a method, in which he embed the numbers in the text based CAPTCHA. It is also called sequenced tagged CAPTCHA. This type of CAPTCHA basically uses two-level of testing. Firstly it matches the character and second it identifies the logical order of the character based on the position of embedded number. As the order has to be identified, it provided additional security.

Burzstein *et. al.* (2009) studied the popular visual CAPTCHA's that can be augmented with the anti-segmentation technique. He applied the systematic evaluation methodology to 15 CAPTCHA's from popular websites and found that 13 are vulnerable

to the automated attacks. He presented series of recommendations for the CAPTCHA designers so that future threats of automated bot attacks can be reduced.

Gossweller *et al.* (2009) proposed a new mechanism for CAPTCHA, which stated that all the complex content of the image should be analysed. The complex content of the image is difficult for the computer-bot to identify but very much easy for the human eye to understand. This method is very much unique because it does not require any text i.e. no text entry for human verification. This method is language independent. The algorithm designed for this mechanism uses image i.e. the image from the repository, usually from the search result. The image selected can be automatically set to the upright orientation. The image is automatically set to the human recognizable upright orientation. Hence, only human will be able to identify, not any computer bot.

Raj *et al.* (2010) presented a paper in which he stated that CAPTCHA based on OCR technique are not safe as they can easily be segmented and thus can be identified by the computer automated bot. While the CAPTCHA based on non-OCR technique are safe, as they utilize the human skill of seeing the image and identifying it. They explain a new concept known as SPC. SPC stands for Sequencing in Picture CAPTCHA. This method is more secure as it provides two level of security. First level is identifying/recognition of object in the images and second level is identifying their logical sequence. SPC generation can be classified into two types- inherent sequencing and non-inherent sequencing.

Edwin *et. al.* (2011) explained in his work that DoS (Denial of Service) attacks are very common. DoS attacks are very harmful as they can occupy resources and produce delay in the network and thus genuine user is not able to do his work because of lack of resources. DoS attacks are also considered as automated network attack. Hence he proposed that, to prevent such type of attack a shield to protect CAPTCHA must be implemented. Text based CAPTCHA i.e. OCR based CAPTCHA are more vulnerable to the attack as there are too many algorithms which can segment the text and CAPTCHA can be broken. If the texts are made too much distorted, there is a possibility that it cannot be broken by the machine, but one disadvantage is that the text cannot be identified by the human also. To overcome this problem a new type of CAPTCHA is introduced i.e. picture based CAPTCHA / graphics based CAPTCHA / image based CAPTCHA. Image based

CAPTCHA's are very much secure from the automated attack, as no malicious program can perform segmentation. No segmentation of image means no identifying of image by computer bot, only human can identify it. The security analysis of image based CAPTCHA has also shown that it is far better than the OCR based CAPTCHA.

Ahmad et. *al.* (2011) reported an automated attack on two widely used CAPTCHA i.e. the CAPTCHA used by GOOGLE and the reCAPTCHA method which was adopted by the GOOGLE. These CAPTCHA were based on anti-segmentation technique i.e. "crowding of characters together" for maximum security. The technique can be applied to the whole text based CAPTCHA family. Ahmad et. *al.* (2011) proposed a new guideline for CAPTCHA design and develops a new framework for more secure CAPTCHA development. Google's reCAPTCHA method uses two distorted texts, one is easily identified by the user and other one is bit tough to identify as these are generally crowding of characters with no meaning and hence bit tough to identify.

Wei-Bin (2012) proposed his work on text based CAPTCHA. He explained that a CAPTCHA is designed in such a way that, it's very hard for the automated computer bot to identify the text. The CAPTCHA is designed such that it contains uppercase letter, lower case letter, special symbol etc. High degree of distortion should be used. The CAPTCHA should not be segmented in any way. With each image a tip is given to the user so that he could identify the text based on that tip. If distortion level is very much high, it is difficult for the computer bot as well as for the human being, to identify the text. So to help the human/user a hint / tip is provided with the distorted image, so that user is able to identify the distorted text. The success rate of human being must be increased for the successful implementation of the CAPTCHA.

Cui *et. al.* (2012) proposed biological motion vision model. It is new type of CAPTCHA model, which is based on moving object identification. After recognizing the moving object, user is able to access the machine or the server. These are the animated CAPTCHA's, which not only check attack against static OCR technology, but also against the moving object. In his research paper, he explained three types of programs i.e. visual recognition program, visual recognition program based on OCR problem, visual recognition program based on non-OCR problem. This type of CAPTCHA is widely accepted. Whenever user make request to the server, server responded to the user with the

picture containing string of random characters and numbers. The user has to identify the sequence of characters to access the server resources.

Saini *et. al.* (2013) deeply examined the working of different types of CAPTCHA. He classified the different types of CAPTCHA and different working conditions of the CAPTCHA's. In their work they gave the guidelines for generating the CAPTCHA and also explained the different application areas of the CAPTCHA. According to them, there are three methods to implement the CAPTCHA i.e. visual method, OCR (Optical Character Recognition) method and non-OCR visual method. And his three basic guidelines for the CAPTCHA were-

a) CAPTCHA should be easy for the human users to identify.

b) CAPTCHA should be easy/ flexible enough to generate (for the machine).

c) It should be hard for the automated computer bot to detect.

Chen (2013) proposed the use of image based CAPTCHA, to differentiate between human and automated computer bots. His method was very much innovative as he proposed the use of characters in the image. He proposed to make characters invisible in the image for the bots. For this, he uses automated image analysis method, which uses scale invariant techniques. It makes easier for the human being to find the location of the embedded characters. This method was capable of reducing the attacks. To test the method 15 users were invited, and the success rate of the method was 86%. If the users who clicked outside the box excluded, success rate becomes 95%. While comparing the logging time with the Google's reCAPTCHA and hello CAPTCHA, this method was faster by 32 seconds and 115 seconds respectively. In this method, random number generation module is used to select the random image and characters. Then colour, size, style and angle verifications are applied to the characters. And finally characters are embedded to the image and final image is generated to test the bot or user.

Chen (2014) explained that safety of the CAPTCHA image depends on its complexity i.e. the complexity of the imaging structure of the image. Therefore he proposed that different recognition method should be applied to the different CAPTCHA's. In the test author worked on detecting numbers between noisy lines and points in

CAPTCHA images. The proposed recognition method achieved an average of 81.05% over 2000 cases.

Ali (2014) presented a paper based on image CAPTCHA. The CAPTCHA was developed using HTML, CSS, and Javascript. The evolutionary prototyping model was used to generate the CAPTCHA. The puzzle based CAPTCHA was generated by the model. The puzzle is solved by the human and hence provides the confirmation to the server and therefore able to distinguish between human and computer bot.

2.2 Summary

Based on the detailed study of the literature, a combination of Automated CAPTCHA and Human Intervention CAPTCHA (HI-CAPTCHA) system is proposed in this research work. The proposed method will not only helps in bot identification, but also solve the real time problems such as in e-polling system, authenticating genuine user in the examination system, attendance system etc. .

The proposed system has Automated CAPTCHA, to filter online bots. The Automated CAPTCHA is based on randomly generated character sets. The length of character set is also random. The user has to answer the questions based on that character set. Questions are generated randomly from the pool of questions. The random generation of question and character set makes it difficult for the bot to identify the text. And hence the system remains secure against the bot attack.

The idea behind the HI-CAPTCHA system is that all the human effort required to solve the Captcha is utilized in some creative way. Besides detecting the bot, the proposed system can be used for e-polling, maintaining records, getting details of the user etc. HI-CAPTCHA system is similar to the Google's reCAPTCHA. As Google's reCAPTCHA performs multiple tasks i.e. identify the bot and solve AI problems. In the same way, the proposed HI-CAPTCHA system is designed i.e. to utilize the human effort. The proposed HI-CAPTCHA will not only differentiate between a bot and human but will also differentiate genuine and invalid users/humans. A genuine user is the one who is authorized and entitled to use the system. For example, in a mobile based classroom exam, a user with the mobile device sitting inside the class is a genuine user whereas the one who is answering from outside the class is an invalid user.

Chapter 3 MATERIALS AND METHODS

CAPTCHA is the unique server side program, which is used to identify the genuine user and it distinguish between computerized automated bots and humans. Thus CAPTCHA security mechanism helps in resource management of the server and makes sure that a resource is occupied by the human user not by computerized bots. In CAPTCHA, a distorted text or image or text embedded in the image is sent to the user screen. The CAPTCHA is formulated such that only human being is able to answer the question and bot is not able to identify the image.

3.1 Problem Formulation

Initially CAPTCHA's were used only for bot identification. But as the AI is evolving CAPTCHA's are becoming smarter and perform multiple tasks. The human effort used to solve the CAPTCHA is used in many creative ways. The machine learning is one of the very important tasks in AI. Often machine learning requires the human effort for learning purpose. For example the Google's famous reCAPTCHA mechanism which uses the human efforts used to solve the CAPTCHA in many ways. For example Google Street mapping helps in learning of different unidentified streets to the machine, Google's book digitalizing program i.e. OCR method is used for the process and all the unidentified words are learned through CAPTCHA mechanism.

CAPTCHA uses a single parameter for the bot identification purpose i.e. distortion level in the text or image. Excessive distortion level is applied to prevent the bot attack, but this can cause difficulty for the human to recognize it. Therefore a balance is kept in distortion level. If less distortion is applied to the image, it can be segmented and easily recognized by the OCR. Therefore beside distortion, many other parameters are used in CAPTCHA. Some of the other parameters are stretching of the text, inverting or left or right orientation of the text, kerning (increasing and removing space between texts), varying fonts, size and style of text etc.

Earlier CAPTCHA's used only for bot identification, but nowadays the human effort, required to solve the CAPTCHA is utilized in creative way i.e. for machine learning process, book digitalization process, making datasets for machine learning etc.

Therefore, the research aim was to utilize the human effort, required for solving CAPTCHA without compromising its security. For this, the detailed analysis of the CAPTCHA mechanism was done and combination of Automated CAPTCHA and Human Intervened CAPTCHA system is proposed. Automated CAPTCHA is used for the identification of online bots. And Human Intervened CAPTCHA is used for authenticating the genuine users. For example a genuine user is the one, who is authorized to use the system.

3.2 Research Methodology

Research Methodology aims to provide the systematic approach in research. It not only helps in problem formulation but also helps in generating step by step solution of the problem. It is used for resolving problems in systematic way and implementation of the research in better way. The research activities that are followed consisted of the phases depicted in the figure 3.1.

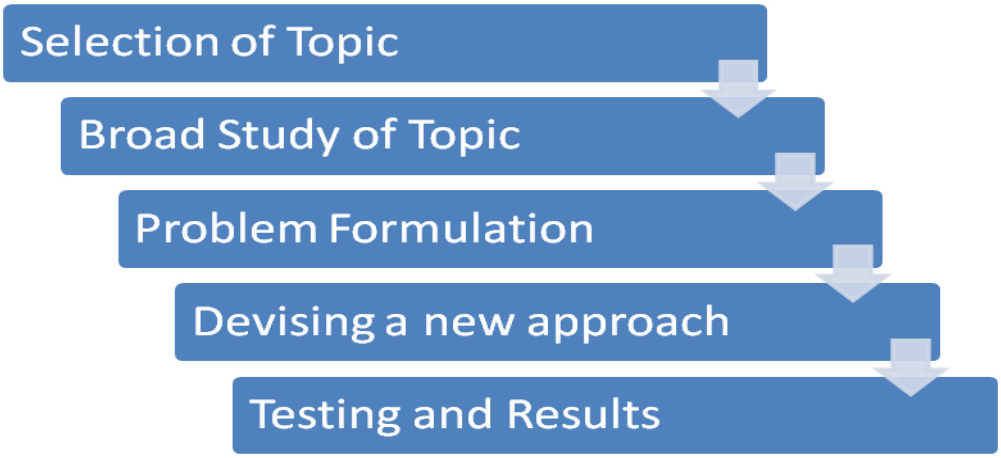


Figure 3.1: Research Methodology

3.3 Tools Used

Following are the main tools which were used for the development of the proposed Human Intervened CAPTCHA system:

a) Spring Tool Suite (STS) for simulation of the web services.

b) Android Studio 2.0 for the application development.

c) MySQL for the database purpose.

d) Virtual Router for local connection.

e) Java and Android to develop the logic.

f) PHP for web based application.

3.4 Installation and Configuration Steps

3.4.1 Android Application

a) Installation of Android Studio 2.0

b) Installation of STS for web services

c) Installation of MySQL Workbench, if database is used in the application

d) Configuring STS for providing local connection.

e) Configuring virtual router connections

3.4.2 Web based application

a) Web Server i.e. XAMPP, WAMP, Apache etc.

b) PHP and JAVA language to develop code.

c) Adobe Dreamweaver for editing the code.

d) MySQL for the Database.

3.4.3 Installation and Configuration Steps for Windows Web Service

a) The first step to see if Wamp Server is successfully installed is to open any Web browser. Mozilla Firefox or any Web browser can be chosen.

b) Wamp Server program must be also started at this step.

c) The address http://localhost/ is to be entered into the Web browser.

d) To test if PHP was installed properly enter the address

e) http://localhost/php_info.php into the Web browser

3.4.4 Pivotal tc Server

Spring Tool Suite is the developer edition of the Pivotal tc server. The Pivotal tc server is the replacement of apache server, as it can run customized spring applications. With its spring insight console, tc server developer edition provides a graphical real time view of application performance metrics and identify and diagnose problems from desktop.

3.4.5 Spring Tool Suite (STS)

The Spring Tool Suite is an eclipse based development environment. STS is customized for developing spring applications. It provides a ready to use frame work for development, debug, run and deploy spring applications. STS also provide integrations for Pivotal Tc Server, Git, Maven, and AspectJ and comes with latest Eclipse release. The spring framework is an application framework that can be used by any java application. It also provides an extension for building web applications.

The Spring Tool Suite supports applications targeting to local, virtual and cloud based servers. It is fully open source and freely available for application development.

3.4.6 Android

Android is the software stack for mobile devices. It is open source and developed by Google. It is based on Linux Kernel. Android is basically designed for touch screen devices such as mobiles and tablets. Android's user interface corresponds to the touch gestures such as tapping of the touch screen, pinching of the screen, swiping of the screen etc. Initially Android was developed by Android, Inc, which later was bought by Google in 2005. Android is very user friendly platform, based on java programming language for the development of the android applications.

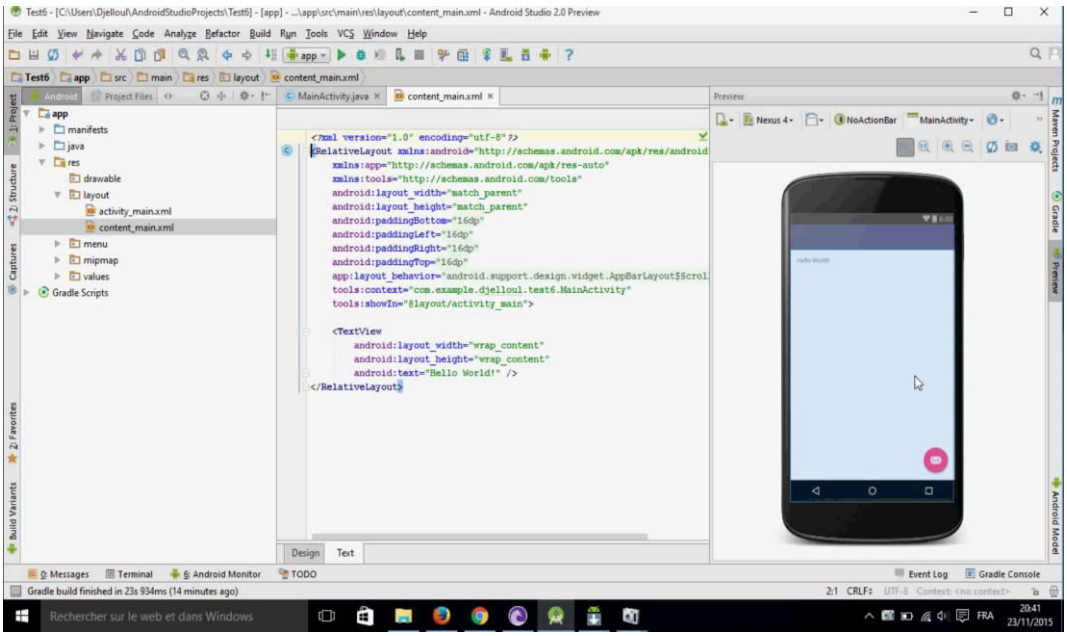


Figure 3.2: Android Studio IDE

Android Studio is the framework or the integrated development environment for the development of Android applications. Android Studio is freely available under the Apache License 2.0. Android Studio is designed specifically for the android application development. It is freely available for Windows, Mac OS X and Linux. It replaced Eclipse Android Development Tools, Google's ADT which was used for the development of android applications. Figure 3.2 shows the Android Studio IDE.

Following table contains all directories and files that Android project must contain:

Table 3.1: Directories/ Files of Android Project

Sl.No.	Dir/Files	Description
1.	**src**	This contains the **.java** source file of the project. It includes **MainActivity.java** source file having an activity class that runs when an app is launched using app icon
2.	**gen**	This directory contains the **.R** file. It is the compiler generated file that references all the resources found in the project. One should not modify this file.
3.	**bin**	This folder contains the Android package file **.apk** built by the ADT during the build process.
4.	**res/drawable-hdpi**	This directory is for drawable objects that are designed for high density screens.
5.	**res/layout**	This is a directory for files that define app's user interface
6.	**res/values**	This is a directory for other various XML files that contain a collection of resources, such as strings and colours definitions.
7.	**AndroidManifest.xml**	This is the manifest file which describes the fundamental characteristics of the app and defines each of its components.

3.4.7 PHP

PHP stands for Hypertext Pre-processor. It is server side scripting language that was designed to develop the web pages i.e. dynamic web pages. Though developed for website designing, it is also used for general purpose programming language. PHP is used for developing dynamic websites or interactive websites. PHP was created in 1994 by Rasmus Lerdorf, to keep the record of who was seeing his online resume. PHP was initially called as Personal Home Page Tools. In mid of 1997, PHP engine was redesigned by Andi Gutmans and Zeev Suraski. They programmed some of the most popular modules of PHP. At that time PHP was popular and was powering many dynamic websites. The website of PHP was php.net. It was run by computer science community.

The following features make PHP a powerful and extensible framework.

a) Less coding

b) Time saving

c) Multiple Language Support

d) Cross Language Inheritance

e) Windows Compatibility

3.4.8 Wamp Server

Wamp Server is a server that can run on any machine. WAMP stands for Windows, Apache, MySQL, and PHP. It is very much suitable for windows platform. It can run on any windows operating system. WAMP 2.1 includes Apache 2.2.17, PHP 5.3.5. Wamp Server is a Windows web development environment. It allows creating web applications with Apache2, PHP and a MySQL database. Wamp Server is a free to use open source project that installs and configures Apache, MySQL, and PHP on a Windows computer. With Wamp Server there is no need to install the individual software components separately.

3.4.9 Apache, PHP and MySQL

a) Apache is a free open source Web server. Apache has been the most popular web server on the Internet since April 1996.

b) PHP is a free open source scripting language designed for producing dynamic, database-driven web pages. PHP is a recursive acronym for "PHP: Hypertext Pre-processor".

c) MySQL is a free open source relational database management system (RDBMS).

3.5 System Requirements

Human Intervened CAPTCHA system can be developed for both i.e. for Android application and for web based application. For the development and implementation of the Android based human intervened CAPTCHA system, the basic system requirements are as follows:

a) Processor : Intel(R)Core™ i3 , 2.00 GHz, 5^{th} gen

b) Installed memory (RAM) : 4 GB

c) System Type : 64-bit operating system, x64- based processor

d) Hard disk : 1 TB

3.6 Proposed Work Description

CAPTCHA is one of the simple mechanisms which helps in bot identification. Today CAPTCHA's are very much secure. All the efforts have been made to make CAPTCHA secure. Many websites use CAPTCHA for identifying the genuine user. Several users interact with these websites and try to solve the CAPTCHA logic. Thus a lot of human effort is involved to solve the CAPTCHA. This human effort may be utilized to solve the CAPTCHA in an effective way. For example Google's reCAPTCHA is used to identify bot and simultaneously it is also used for machine learning i.e. learning of Google's Street Map, learning of characters of the book etc. An Automated CAPTCHA system is implemented and a new modified CAPTCHA i.e. human intervened CAPTCHA (HI-CAPTCHA) system proposed in this work.

The Automated CAPTCHA is used for the detection of online bots. It uses randomly generated character set and random questions based on the character set to identify the bot. Questions are selected randomly from the pool of questions. Too much randomness makes it safe against the bot attack.

The proposed HI-CAPTCHA not only differentiates between a bot and human but also differentiate genuine and invalid users/humans. A genuine user is the one who is authorized and entitled to use the system. For example in a mobile based classroom exam, a user with the mobile device sitting inside the class is a genuine user whereas the one who is answering from outside the class is an invalid user. Hence, the users answer to the proposed HI- CAPTCHA determines whether he is genuine or invalid. Thus user's effort to answer HI-CAPTCHA is utilized in an effective manner.

In the proposed HI-CAPTCHA system questions are generated by the admin. The generated questions are send to the users/clients. The answer of the question is varied as per the admins requirement. For example in question, What is the colour of the Admin shirt? The answer of the question is not fixed as the admin can wear any colour shirt in any day. The answer depends on the admin shirt. Similarly in the question, Is admin moving or standing? The answer changes according to admin's direction of movement. The system works with local server, where the numbers of users connected to the local server are fixed. The HI-CAPTCHA questions not only detect the bots but also help in authenticating the genuine user in selective mobile based system. For example, e-polling results, attendance based systems, online exam etc.

3.6.1 Block diagram of Proposed System

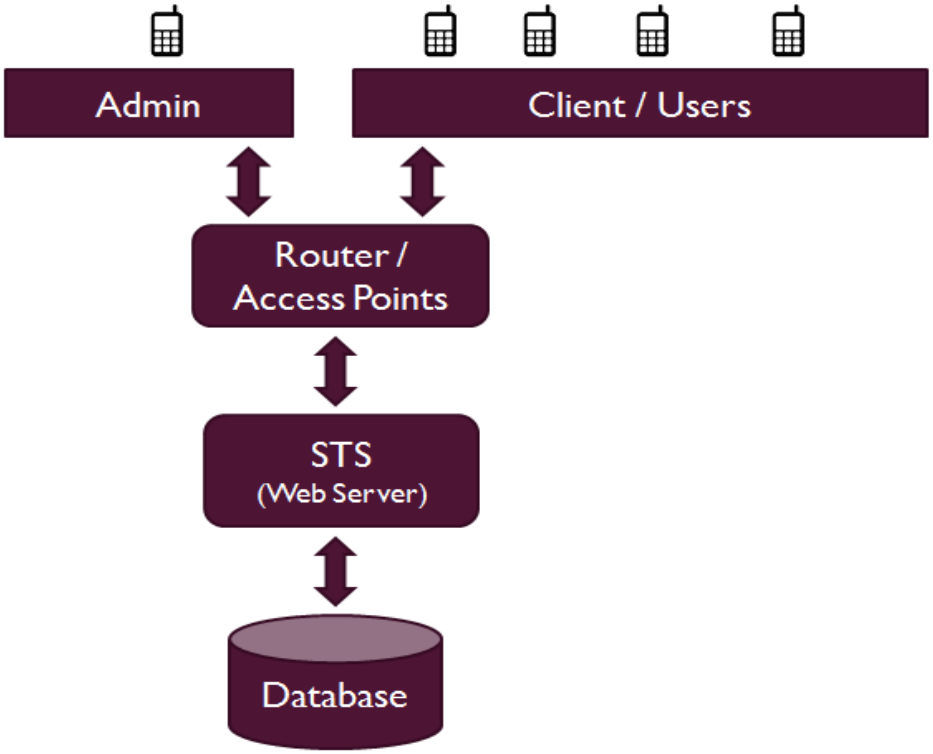


Figure 3.3: Block diagram of Human Intervened System

3.6.2 Features of Proposed System

The proposed system is very much efficient in bot detection. It maintains the security level of CAPTCHA system. The proposed CAPTCHA system is designed in such a way that, it can perform multiple tasks. This CAPTCHA is very much similar to the Google's reCAPTCHA in form of its working. As Google's reCAPTCHA helps in machine learning, this CAPTCHA helps in generating the result of the common question. Thus this CAPTCHA is called the Human-Intervened CAPTCHA system.

a) All the questions are generated by the Human and the answer of these questions are not fixed. The answer depends upon the present situation, present time and present location of the admin. The answer of the questions is variable. Therefore, no automated computerized bot is able to predict it and answer it. Thus the system remains secure against the bot attack.

b) As the CAPTCHA is human-intervened, the questions generated by the human can be customized in such a way that, it can be used to answer the multiple

queries, such as how many students are present in the classroom? E-voting in a room or hall etc.

c) Ability to control the questions and answers, gives very much flexibility against the bot.

d) Ability to select the random questions by the admin makes it hard for the bots to answer the question and hence bot attack can be neutralized.

3.6.3 Parameters and Their Usage

a) Human Intervened CAPTCHA

The proposed system is the Human intervened CAPTCHA system. In this system questions are generated by the admin and answer of the question is not fixed. Based on situation, time and location, the answer of the question changes. Hence bots are not able to detect the answer and the system remains secure against bot attack. The proposed system is designed in such a way that the human effort required to solve the CAPTCHA is utilized in creative way i.e. answering some of the backend questions, getting details of the users. The proposed system is very much effective in local server environment.

b) Questions

Human intervened CAPTCHA system has customized questions. The questions are generated by the admin. The questions are designed in such a way that, the answer always varies. The answer always depends on current time, situation and location. The variable questions provide additional security against bot attack. Some of such questions are as follows-

- What is the colour of the shirt admin wearing?
- Is admin moving or standing?
- Which direction admin is standing?

c) Answer of the Questions

The proposed system has fixed questions. All the questions of the system are generated by the admin. The admin generates the question in such a way that, its answer is not same. Its answer depends on current time, position and situation. For identifying bots in small areas and for performing multiple tasks, this system can be used. The answers of the question such as:

- Is admin moving or standing?

  Probable answer is either moving or standing.

  The answer of the question varies according to the movement of the admin.
- What is the colour of the shirt admin wearing?

  Probable answer is red, green, blue, black etc. The answer of the question depends on the colour of the shirt admin wearing.
- Which direction admin is standing?

  Probable answer is east or west or south or north. The answer of the question varies according to the location, time and situation of the admin.

d) Time to solve the CAPTCHA

Human intervened CAPTCHA system requires some time to solve by the clients/users. As the answer of the questions is not fixed, the clients/users answer depends on admin. The proposed system gives additional power to the admin i.e. the power of deciding the time required to solve the question by the user. With each question that is send to the user, the admin also sends the time within which the answer must be submitted. This amount of time also restricts the user from attempting all the answers. Hence, this feature restricts the unauthorized access to the system. The time given by the admin to solve question depends upon the complexity level of the question. The time can be as-

- 2 min
- 4 min
- 6 min
- 8 min etc.

e) Record Maintenance

The proposed system is multi-tasking in nature. Its working is similar to the Google's reCAPTCHA. Google's reCAPTCHA not only helps in bot detection, but also helps in machine learning. For example, **(1)** In Google's program of learning books, reCAPTCHA helps in recognizing new unidentified words. **(2)** In Google's street mapping reCAPTCHA helps in learning of new street. In the same way, in the proposed system the human efforts required to solve the CAPTCHA is utilized in creative way to answer the multiple backend questions like knowing the details of the users, e-polling results, counting of the users etc. In this system, as the user

submits the answer, all its details are saved to the database. Thus we can answer the questions, besides detecting the bots. Some of the backend tasks that can be performed by the proposed system are as follows:

- How many users are present / connected to the server?
- Details of the connected users i.e. ID, Name, and Course etc.
- E-voting in a room or in small area.
- Bot detection

f) Automated CAPTCHA

The proposed system is designed in such a way that besides sending the human intervened CAPTCHA, it can also send the automated CAPTCHA. The sole aim of automated CAPTCHA is to identify the automated bots only. The automated CAPTCHA consists of combination of characters, numbers and special symbols. The characters can be in upper case or in lower case. The user has to answer different types of the questions based on the character set. The questions are designed in such a way that only genuine user is able to answer the question and hence automated bot attack can be stopped. The distortion level may be applied to the character set, to provide additional security to the system. But random generation of the questions makes it very much difficult for the automated bots to answer the question.

g) Automated CAPTCHA Questions

The primary aim of automated CAPTCHA is to detect automated bots and stop them from entering into the system. The automated CAPTCHA consist of character sets i.e. the combination of upper case and lower case characters, numbers and special symbols. This character set is send to the user and a question based on that character set is send. The questions are designed such that it is very hard for the bots to answer. Hence bot attack can be stopped. The questions are picked randomly from the pool of questions. Some of the automated CAPTCHA questions are as follows:

- Identify the alphabets in the text?

  Based on this question an algorithm is designed to answer it. If this question comes in the CAPTCHA, all the alphabets are identified and checked.

- Identify the numbers in the text?
  An algorithm is designed to find the numbers in the character set. If user identifies it, he passes the CAPTCHA test.
- What is the colour of the text?
  In this question, user has to answer the colour of the character set.
- Solve the equation?
  User has to solve the equation, as a part of CAPTCHA test. If user solves it, he is genuine user, if not it is an automated bot.
- Submit the nth character of the word?
  Based on this question an algorithm is designed to answer it. If this question comes in the CAPTCHA, nth character has to be identified and checked.

h) Automated CAPTCHA Questions difficulty level

To provide the additional security, the admin can even apply the distortion level to the character set generated. But only limited distortion level is applied to the character set, so that it can be easily answered by the user.

3.6.4 CAPTCHA Time to Solve

It is the time that is required to solve the CAPTCHA. The time is in seconds i.e. from when the code was created, to when it was solved. The simple time function is as follows:

```
public function getTimeToSolve ()
{
        return $this->_timeToSolve;
}
```

HI-CAPTCHA system requires some time to solve by the clients/users. It is the maximum amount of time given by the admin to the user/client to solve the CAPTCHA. With each question that is send to the user/client, the admin also sends the time within which the answer must be submitted. This amount of time also restricts the user from attempting all the answers. Hence, this feature restricts the unauthorized access to the system. The time given by the admin to solve the CAPTCHA question depends upon the complexity level of the question.

Chapter 4 RESULTS AND DISCUSSION

The outcome and results obtained by the proposed system are briefly explained in this chapter. This chapter also discusses about the failure region of the proposed method. The proposed Human Intervened CAPTCHA (HI-CAPTCHA) method is implemented with client and admin modules. The proposed system is developed on Android platform. The main software's used for the development of the proposed system are Pivotal tc server, Spring Tool Suite, Android 2.0. Developed Mobile based system includes, Admin module used by admin for the selection of CAPTCHA, Client module used by client for answering the CAPTCHA questions and Web server for deploying admin module and client module.

Result and Discussion section includes 1) Generated CAPTCHA, 2) Selection of CAPTCHA, 3) Client view of Automated CAPTCHA, 4) Client view of HI-CAPTCHA, 4) HI-CAPTCHA test case, for identifying the genuine user and 6) HI-CAPTCHA test case, if user inside class not connected to server.

4.1 Generated CAPTCHA

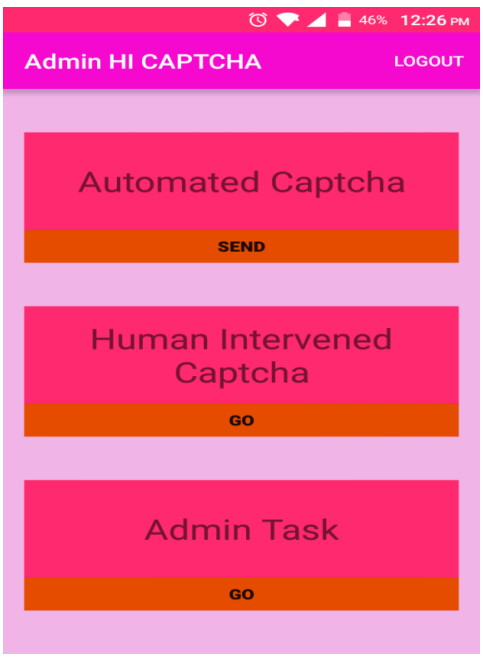
This section gives the detailed explanation of the proposed CAPTCHA system. All the challenges that come after applying the CAPTCHA system are described here. The developed system consists of two type of CAPTCHAs one is automated CAPTCHA and second is Human Intervened CAPTCHA (HI-CAPTCHA). Figure 4.1 shows the selection of two types of CAPTCHA.

4.1.1 Automated CAPTCHA System

The developed system includes has automated CAPTCHA for detecting online bots. The primary aim of automated CAPTCHA is to detect automated bots and stop them from entering into the system. Automated CAPTCHA depends on the character sets that are generated automatically. The character set is generated randomly. It consists of special symbol, upper case letter, lower case letter, numbers etc. The random generation of the character set makes it difficult for automated bots to recognize it. The length of the automatic character set is also random, which makes it even more difficult for the bots. Hence the computerized bots are stopped from entering into the system.

4.1.2 Human Intervened CAPTCHA (HI-CAPTCHA)

The developed system includes proposed new modified CAPTCHA i.e. HI-CAPTCHA as second component. In this proposed HI-CAPTCHA system questions are generated by the admin and answer of the question is not fixed. Based on situation, time and location, the answer of the question changes. Hence bots are not able to detect the answer and the system remains secure against bot attack. Today CAPTCHA are designed in such a way that, the human effort required to solve the CAPTCHA is utilized in many ways. The proposed system is also designed in such a way that the human effort required to solve the CAPTCHA is utilized in creative way i.e. answering some of the backend questions, getting details of the users. The proposed HI-CAPTCHA is designed to differentiate between bot and human. It will also differentiate between genuine user and invalid user/human. A genuine user is the one, who is authorized to use the system. Figure 4.2 shows the HI-Captcha.



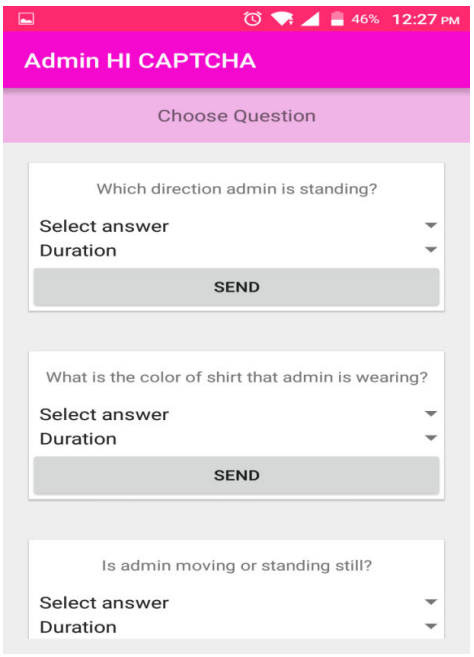


Fig 4.1: Selection of Captcha **Fig 4.2: Selection of HI-CAPTCHA Question**

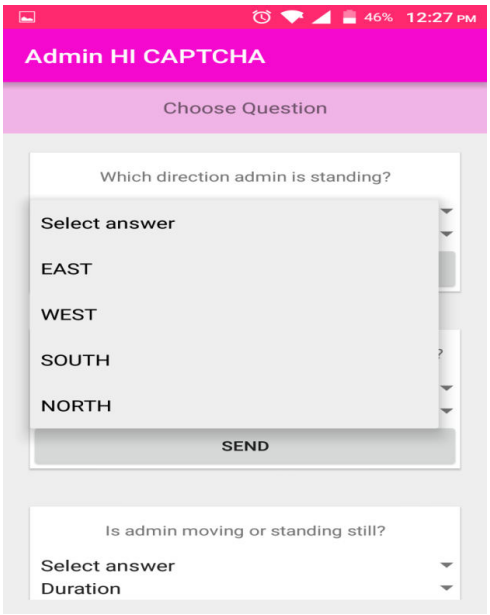


Fig 4.3: Selection of answer

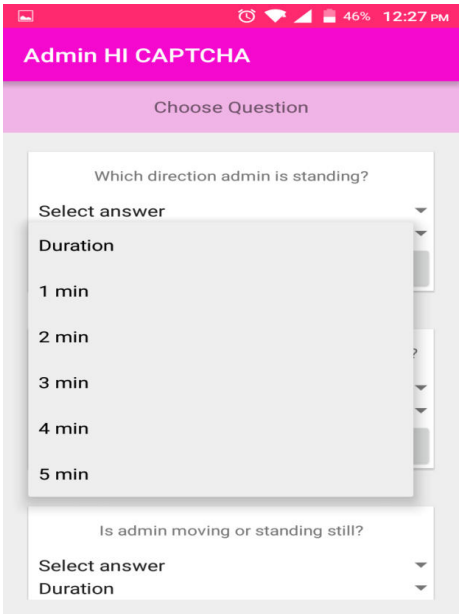


Fig 4.4: Selection of Time

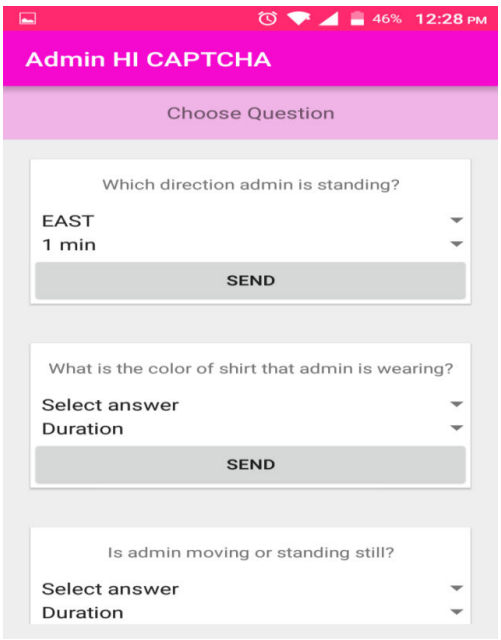


Fig 4.5: Sending selected question

4.2 Selection of CAPTCHA

The developed system consists of two types of CAPTCHAs i.e. Automated CAPTCHA and Human Intervened CAPTCHA. Based on requirement, admin can select any of the CAPTCHA from the two. The selected CAPTCHA will go to the client for authentication. The purpose of both the CAPTCHA is different. Based on the situation and

requirement of admin, the CAPTCHA's are selected. Figure 4.1 shows the Admins screen and the tasks that Admin can perform.

4.2.1 Automated CAPTCHA

The proposed system has Automated CAPTCHA for the identification of online bots. Bot or Spam is the automated computer programs that enter into the system and occupy resources or spy into system. These are very harmful, as they create serious security problems. The Automated Captcha is designed to stop the bot from entering into the system. The automated questions are designed in such a way that only human can answer the question not any computerized bot. The questions are generated randomly.

The proposed automated Captcha system contains the pool of questions. The questions are selected randomly from the pool. The Automated Captcha questions depend on randomly generated character sets. Too much randomness in the question generation provides additional security against the bot. Hence the system remains secure against the bot attack.

4.2.2 Time to solve Automated CAPTCHA Question

Automated Captcha is designed to identify the bots only. Hence the time factor is also important. The minimum amount of time given to solve the automated Captcha question is 2 min only. Time plays very important role in automated Captcha because:

- Time prevents bots from applying brute force technique to solve the question.
- Time reduces the waiting of the admin.

4.2.3 Human Intervened CAPTCHA

The major research area of this thesis is the Human Intervened CAPTCHA system, where the human can intervene in-between. For example, in a classroom a teacher can intervene in-between the validation of the answer and attendance of the student. Main aim of HI-CAPTCHA system is detect the bot or an unauthorized human and stop them from entering into or operate with the system. In this system, admin has the power to generate the question and the answer of the question is not fixed. The answer depends upon location, time and situation of the admin. As the answers are not fixed, no computerized bot would be able to detect. Hence bots are stopped. The HI-CAPTCHA is designed in such a way that, it will not only detect the bots but will also differentiate between genuine user and invalid user. A genuine user is the one who is authorized to use the system, while

the invalid user/human is the one who is answering from outside the class. The idea behind HI-CAPTCHA is that, human effort required to solve the CAPTCHA is utilized in an effective way. Thus, the HI-CAPTCHA can be utilized in very much effective in e-polling, conducting small exam, getting details of the users etc. The proposed Human Intervention CAPTCHA system works very much fine with the local server.

4.2.4 Selection of Human Intervened CAPTCHA Question

The Human Intervened CAPTCHA system consists of many questions. The admin has the power to select any question from the list. The admin has also the power of associating time to the question i.e. the maximum amount of time within which client or user has to answer it. The answer of the question is not static i.e. answer changes as the situation changes, time changes or the location changes. Figure 4.2 shows the questions available with the admin. The admin is supposed to select the one question from the list. Then, depending upon situation, he selects the correct answer. The question and its corresponding answer get saved in the database for further verification of the clients.

4.2.5 Time to solve Human Intervened CAPTCHA Question

The proposed Human Intervened CAPTCHA system is designed in such a way that the maximum time required to solve the CAPTCHA questions is determined by the admin. The system gives power to the admin to decide the time i.e. 2 min, 4 min, 6 min, 8 min etc. Depending upon the complexity of the CAPTCHA question, the time is decided by the admin. It is decided in such a way that all the users can answer the question within the time limit. Figure 4.4 shows the selection of time.

The time factor is very much crucial in the proposed system because of the following reason:

- Based on the complexity of the question, the time is decided by the admin. It is the maximum amount of time given by the admin to the user to answer the question.
- The time factor given in the proposed system helps in reducing un-necessary waiting of the admin. I.e. admin has to wait only for the time allotted to the question.

4.3 Client View of the Automated CAPTCHA

The clients are supposed to sign up (Figure 4.6 shows the signup page). The automated Captcha is designed to stop the bots. It consists of the randomly generated questions. The questions are designed such that only answered by the human. The random generation of the character set makes it difficult for automated bots to recognize it. The length of the automatic character set generation is also random, which makes it even more difficult for the bots. Hence, the computerized bots are stopped from entering into the system. Figure 4.7 shows the Client view of the automated questions.

Some of the automated CAPTCHA questions are as follows:

- What are the alphabets in the text?

  Based on this question an algorithm is designed to answer it. If this question comes in the CAPTCHA, all the alphabets are identified and checked.
- What are the numbers in the text?

  An algorithm is designed to find the numbers in the character set. If user identifies it, he passes the CAPTCHA test. s
- What is the colour of the text?

  In this question, user has to answer the colour of the character set.
- Solve the equation?

  User has to solve the equation, as a part of CAPTCHA test. If user solves it, he is genuine user, not any automated bot.
- Submit the nth character of the word?

  Based on this question an algorithm is designed to answer it. If this question comes in the CAPTCHA, nth character has to be identified and checked.

4.4 Client View of the Human Intervened CAPTCHA

Client has to answer the human intervened questions. The time for answering the question is mentioned in the question. The user has to answer the question in fixed amount of the time. The answer of the question designed in the human intervened Captcha system is not fixed. The answer can vary day to day. For Example:

- If the question is what is the colour of shirt, admin wearing?

  The answer of the question varies day to day. Admin can wear different colour of shirt. So the user has to see the admin and then select the answer from the

drop box. If the user is genuine i.e. present in that particular area, he will answer it correctly, otherwise not.

- In admin moving or standing?

  Answer depends on the location, time, situation and the wish of the admin.

- In which direction admin is moving?

  The answer of the question may be East, West, North or South. The User has to answer the question by seeing the admin. The answer varies, according to location, time and situation.

The aim of designing the HI- CAPTCHA system is utilizing all the human effort required to solve the Captcha in some creative way. The questions are designed locally and could be answer in the local server. Human Intervened Captcha system is very much effective in local server, because designed questions can be answered by seeing the admin. The answer of the designed question is not fixed. Hence computer bots are not able to answer and the system remains secure against the bot attack. Figure 4.8 shows the Client view of the Human Intervened Captcha questions.

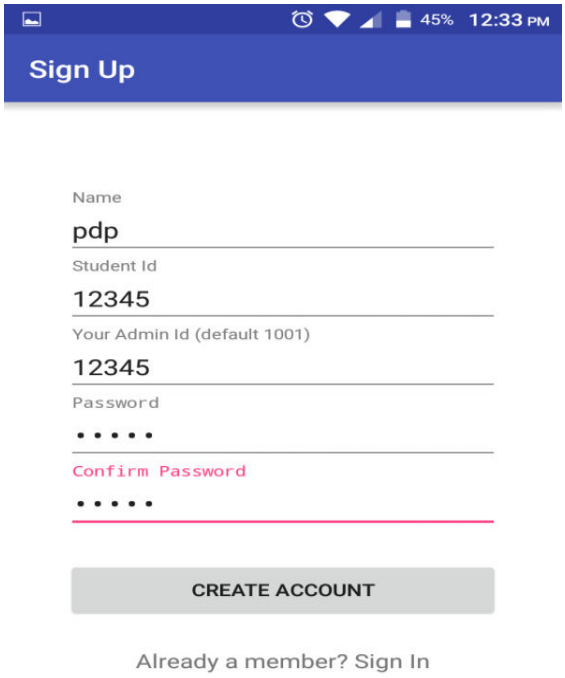


Fig 4.6 Client Sign Up Page

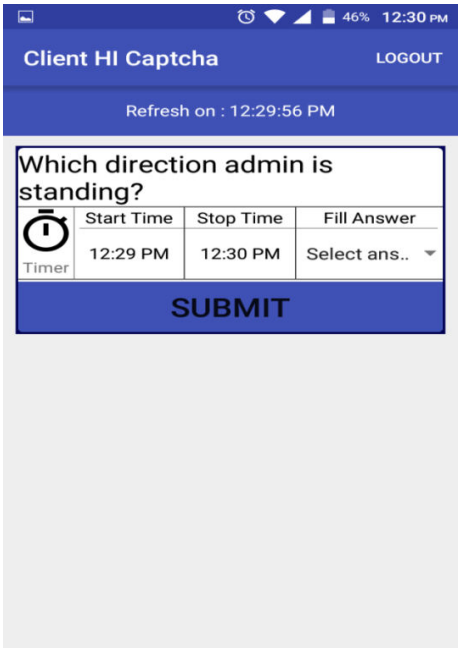


Fig 4.7 Client Automated Question

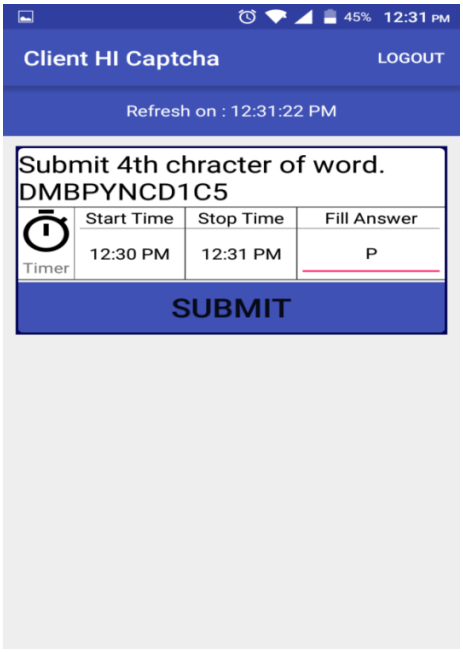


Fig 4.8 Client HI Question

4.5 Human Intervened CAPTCHA (HI-CAPTCHA) Test Cases for Identifying the Genuine User

HI-CAPTCHA is used for the identification of the genuine users. A genuine user in the system is the one, who is authorized to use the system. For example, in the mobile based online examination system, all the users inside the class are genuine user, while the users outside the class are not genuine.

Following test case is conducted to prove the result of the HI-CAPTCHA.

a) Admin selects the HI-CAPTCHA to authenticate the genuine user.
b) A question selected by admin is, Is admin moving or standing?
c) Answer of the question, based on the current time, situation or location of the admin is standing.
d) Time selected by Admin is 2 min. It is the maximum time, in which user has to answer the question.

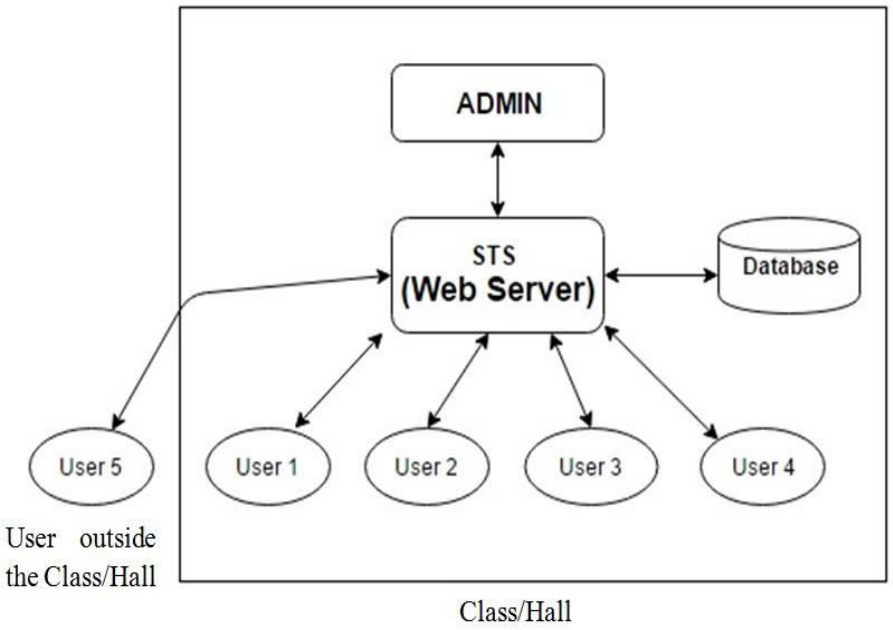


Figure 4.9: Human Intervened CAPTCHA (HI-CAPTCHA) Test

All the users inside the class/hall will answer correctly, as they are seeing the admin. But the users outside the class/ hall are not able to answer correctly, because they are not seeing the movement of the admin. Hence, the admin will be able to identify the genuine users. Figure 4.9 shows the diagrammatical view of the HI-CAPTCHA test case, for identifying the genuine user.

Similarly, the user may select the following question for identifying the genuine user.

a) In HI-CAPTCHA mode, admin can select question, Which Colour shirt, is admin wearing?
b) Admin can wear any colour shirt in any day. So answer of the question depends on the admins shirt colour. Suppose, admin is wearing Red colour shirt. So the answer of the question is Red.
c) Let the time selected by admin to answer the question is 1 min. All the users have to answer within 1 min time limit.
d) Admin sends the selected question to the users, to authenticate the genuine users, who are authorized to use the system.

The users inside the class will answer correctly, because they are seeing the admin and hence know the colour of admin shirt. Therefore, they are the genuine users and authorized to use the system.

4.6 HI-CAPTCHA Test Cases, if user inside class is not connected to server

HI-CAPTCHA is used for identifying genuine user. All the users inside the class are genuine user, as they are authorized to use the proposed system. But the case may arise, when the user inside the class is not connected to the server. In that case, user is not able to utilize the services of the system and he is considered as unauthorized user. But the user inside the class is always genuine user. There are many conditions due to which this condition may arise.

a) Application may be not installed in the device. The user has to install the HI-Client CAPTCHA app.

b) The device is not configured properly to use the system.

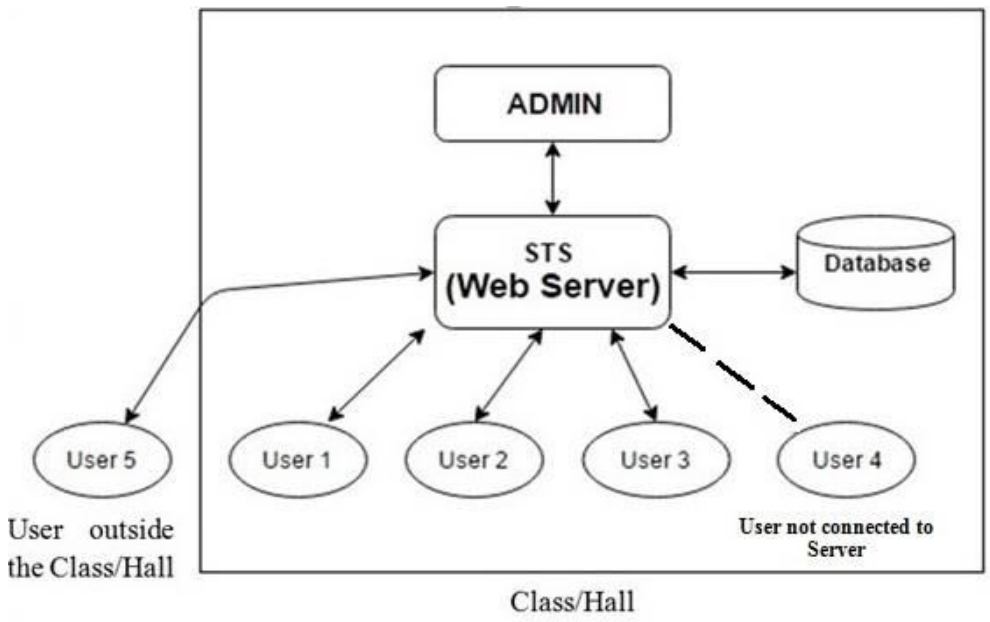


Figure 4.10: HI-CAPTCHA Test case, if user is not connected

If the user is not connected to the server, following condition may be checked:

a) The user has to sign up to the HI-Client CAPTCHA app.

b) The user inside the class/hall has to retry login in the client HI-CAPTCHA application.

c) If he logins successfully, he may continue. But if user fails to login, the user has to check the settings of the mobile i.e. check wifi settings, check IP settings etc.

d) If the user logins successfully and waiting for the CAPTCHA questions. The user has to wait as there might be delay in the network. Further the user may contact to the admin for help.

Chapter 5 SUMMARY AND CONCLUSION

5.1 Summary

This thesis gives the detailed explanation of the proposed CAPTCHA system i.e. Automated CAPTCHA and Human Intervened CAPTCHA (HI-CAPTCHA) system. Detailed literature survey has been performed for the proposed system. All the challenges that come during the implementation of the CAPTCHA system are described in the thesis. The proposed thesis gives the detailed description of two types of CAPTCHA's.

5.1.1 Automated Captcha System

The proposed system has automated CAPTCHA for identification of online bots. A detailed study of the Automated CAPTCHA is done with the aim of detecting automated bots and stopping them from entering into the system. Automated CAPTCHA depends on the randomly generated character sets. It consists of special symbols, upper case letters, lower case letters, numbers etc. The random generation of the character set makes it difficult for automated bots to recognize it. The length of the automatic character set is also random, which makes it even more difficult for the bots. Hence the computerized bots are stopped from entering into the system.

5.1.2 Human Intervened Captcha System

The proposed system is the Human intervened CAPTCHA (HI-CAPTCHA) system. In this system, a question is generated by the admin and answer of question is not fixed. Based on situation, time and location, the answer of the question changes. Hence bots are not able to detect the answer and the system remains secure against bot attack. As the AI is evolving CAPTCHA's are becoming stronger against the bot attacks. Today CAPTCHA are designed in such a way that, the human effort required to solve the CAPTCHA is utilized in many ways. They not only detect the bots, but also answer some of the basic questions of the AI. The proposed HI-CAPTCHA system is also designed in such a way that the human effort required to solve the HI-CAPTCHA is utilized in creative way i.e. to differentiate between genuine user and invalid users.

The proposed HI-CAPTCHA will not only differentiate between a bot and human but will also differentiate genuine and invalid users/humans. A genuine user is the one who

is authorized to use the system. For example, in a mobile based classroom examination system, a user with the mobile device sitting inside the class is a genuine user whereas the one who is answering from outside the class is an invalid user. Hence, the user's answer to the proposed HI-CAPTCHA will determine whether he is genuine or invalid. Thus user's effort to answer the HI-CAPTCHA is utilized in an effective manner.

5.2 Conclusions

CAPTCHA's are very effective way for stopping bots and reducing spams. Captcha keep web data secure from intruders. Almost every website contains Captcha in one form or other. All the Sign in and sign ups or any form submission over the internet contains Captcha. The developed system has an Automated Captcha system for the identification of online bots. The automated CAPTCHA depends on the randomly generated **character sets**. The length of the character set is also random. It consists of uppercase letter, lowercase letter, numbers and special symbols. The questions are generated randomly based on character set. The questions are designed in such a way that, only human can answers them, not any bot. Thus the system remains secure against the online bot.

As Artificial Intelligence is evolving, the need for developing new and advanced type of Captcha's has been arrived. Google's reCAPTCHA is the example of advance Captcha. It not only detects the bot, but also helps in machine learning process. In the same way the proposed HI-CAPTCHA system is also designed for multi-tasking environment. The proposed HI-CAPTCHA will not only differentiate between a bot and human but will also differentiate genuine and invalid users/humans. The proposed HI-CAPTCHA system can be designed for attendance system, e-polling system or for generating the details of the users. It works very fine in the local server environment.

The dual combination of Automated Captcha and Human Intervened Captcha is very much effective against the bot or spam detection. Both the Captcha's in one system not only provides the security against the bots but also helps admin in finding the genuine user or helps in getting the details of the users. The system may be customize in many ways i.e. for e-polling system, attendance system, online examination system etc. In all these systems, the proposed combination of HI-CAPTCHA and the Automated Captcha helps admin in bot detection and finding the genuine users.

5.3 Recommendations for future work

As AI is evolving, there is a lot of advancement in the automated techniques that make it more difficult for the Captcha security. Many automated computer programs have been made, that intrudes or violates the Captcha security. Although a lot of efforts have been made, in designing the system which remains secure against bots in any condition, still in fast moving era, system might break. So following future works are proposed for Captcha security.

a) Interface optimization

The proposed system has to be developed for visually and audibly challenged persons. Text To Speech (TTS) features may be added to the system for the challenged ones.

b) Multiple Language Support

The System may be designed with the multiple language support. As the system runs in local server interface, multiple language support may helps in many features. For example, generating information about crop from a group of farmers, getting any similar type of information from the group of people etc.

c) Distortion Level

Distortion level may be applied to the automated character sets, for making system more secure against the bot attack.

d) Larger scale study

To produce statistically significant results, a larger scale study with truly randomized samples should be carried out.

e) Statistical Data

The proposed system can be designed to get the statistics or 2D information from the group of people. A possibility may be explored in this direction.

LITERATURE CITED

Ahmad, S., Yan, J. and Mohammad, T. 2011. The Robustness of Google CAPTCHAs. *Computing Science, Newcastle University.*

Ahn, L. V., Benjamin, M., Colin, M., David, A., Blum, M. 2009. reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science, 321(5895),1465-1468*

Ahn, L. V., Blum, M. and Langford, J. 2004. Telling Humans and Computers Apart Automatically: How lazy cryptographers do AI. *Communications of the ACM.*

Ahn, L. V., Blum, Nicholas, J. and Langford, J. 2000. The official CAPTCHA website. *www.captcha.net*.

Ahn, L. V., Blum, Nicholas, J., and Langford, J. 2003. Using Hard AI Problems for Security. *Proceedings of Eurocrypt.*

Ahn, L. V., Maurer, B., Colin, M., Crawford, M., Staake, R., Blum, M. reCAPTCHA Project Online. *htttp://www.recaptcha.net.*

Ali, B. 2014. Development of CAPTCHA system based on puzzle. *Computer, Communications, and Control Technology I4CT.*

Azad, S. & Jain, K. 2013. CAPTCHA: Attacks and Weaknesses against OCR Technology. *Global Journal of Computer Science and Technology*, *13*(3)

Bursztein, E., Matthieu, M., Mitchell, J. C. 2011. "Text-based CAPTCHA: Strengths and Weaknesses", *ACM Computer and Communication Security, CSS'2011*

Chen C. and Wu, Z. 2013. Anti-SIFT Images Based CAPTCHA Using Versatile *IEEE International conference on Information Science and Applications (ICISA)*

Chen, C. 2014. A Study on CAPTCHA Recognition Intelligent Information Hiding and Multimedia Signal Processing IIHMSP. *Tenth International Conference on* (pp. 395-398). IEEE

Chew, M. & Henry S. 2003. Baffle Text: a Human Interactive Proof *Fifth International Conference on* (pp. 807-810). IEEE

.

Cui, J., Zhang, W., Mei, J., Wang, L. and Wang, X. 2012. CAPTCHA Design Based on moving Object Recognition Problem. *International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery*

Edwin, R. Devassy, D. and Jagannivas, J. A New Architecture for the Generation of Picture Based CAPTCHA. *Electronics Computer Technology (ICECT), 2011 3rd International Conference on (Vol. 6, pp. 67-71). IEEE*

Gossweller R., Kamvar, M. and Baluja, S. 2009. What"s Up CAPTCHA? A CAPTCHA Based on Image Orientation. *MADRID, pp. 841-850, 2009*

Gupta A., Jain, A. 2009. Sequenced Tagged CAPTCHA: Generation and its Analysis. *Advance Computing Conference. IACC 2009. IEEE International*

Hidalgo, J. M. G., Alvarez, G. CAPTCHAs: An Artificial Intelligence Application to Web Security. *Advances in Computers*, *83*(1), 109-181

Hruska, J. 2008. Right Back at Ya: CAPTCHA: Bad Guys Crack Gmail, Hotmail. *http://arstechnica.com/security/2008/04/gonein60secondsspambotcrackslivehotmail captcha*

Kluever, K. A. 2008. Breaking the PayPal HIP: A Comparison of Classifiers *Document and Pattern Recognition Lab, Department of Computer Science, Rochester Institute of Technology.*

Kulkarni, S., Fadewar, H. S. 2013. CAPTCHA Based Web Security: An Overview. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(11)

Kumar, C., Kevin, L., Patrice, S. & Czerwinski, M. Designing Human Friendly Human Interaction Proofs HIP. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 711-720). ACM

Kumar, M., Dhir, R. 2013. Design and Comparison of Advanced Color based Image CAPTCHAs. *International Journal of Computer Applications*,*61*(15), 24-29

Mark, D., Abadi, M., Bharat, K., Andrei Z. 1998. Patent US 6195698 - Method for selectively restricting access to computer systems - *Google Patents*,

Mori, G. and Malik, J. 2000. Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA *Proceedings. 2003 IEEE Computer Society Conference on (Vol. 1, pp. I-134). IEEE*

Motoyama, M., Levchenko, K., Kanich, C., McCoy, D., Geoffrey M. Re:CAPTCHAs–Understanding CAPTCHA-Solving Services in an Economic Context. *USENIX Security Symposium (Vol. 10, p. 3)*

Oleg, S., Claudia C., Fernando, U., Vicente, A. 2014. Pattern Recognition- Breaking text-based CAPTCHAs with variable word and character orientation. *Pattern Recognition*, *48*(4)

Patrick J. & Robert, O. 2011. Sentence First CAPTCHA: Proposal and Study of a textbased CAPTCHA Scheme

Phoenix, D. S. 2013. Vicarious AI passes first Turing Test: CAPTCHA *http://vicarious.com*

Rahman, R. 2012. Survey on CAPTCHA Systems. *Journal of Global Research in Computer Science*

Raj, A., Pahwa, T. and Jain, A. 2010. Picture CAPTCHAs with Sequencing: Their Types and Analysis *International Journal of Digital Society, vol. 1, no. 3, pp. 208-220*

Ryan F., Luu G. and McMohan, P. Learning to Read Obscured and Distorted Text in Images. *CS 229 Project Report: Cracking CAPTCHAs.*

Saini, B. S. and Bala, A. 2013. A Review of Bot Protection using CAPTCHA for Web Security. *IOSR Journal of Computer Engineering IOSR-JCE.*

Sarika, 2013. Understanding CAPTCHA: Text and Audio Based CAPTCHA with its Applications. *International Journal of Advanced Research in Computer Science and Software Engineering*

Schow, C. M. 2011. Report on Advances in the Field of Artificial Intelligence Attributed to CAPTCHA.

Singh, V.P. and Pal, P. 2014. Survey of Different Types of CAPTCHA *IJCSIT. International Journal of Computer Science and Information Technologies, Vol. 5*

Verma, R. and Kaur, R. 2014. Enhanced Character Recognition Using Surf Feature and Neural Network Technique

Waseem, S. 2008. POSH: A Generalized CAPTCHA with Security Applications. *Proceedings of the 1st ACM workshop on Workshop on AISec*

Wei-Bin, L. 2012. A CAPTCHA with Tips Related to Alphabets Upper or Lower Case *Broadband, Wireless Computing, Communication and Applications BWCCA.*

Yan, J. and Ahmad, A. S. 2008. A Low-cost Attack on a Microsoft CAPTCHA *School of Computing Science, Newcastle University, UK*

VITA

The author, Pradeep Giri, was born on 18th April 1991 in Pithoragarh district of Uttarakhand. He passed his High School and Intermediate Examination from Maharishi Vidya Mandir, Pithoragarh (affiliated to C.B.S.E Board) in 2006 and 2008 respectively. He earned his bachelor's degree in Computer Science and Engineering from Dev Bhoomi Institute of Technology, Dehradun (affiliated to Uttarakhand Technical University) in 2012. He took admission in the College of Post Graduate Studies at Govind Ballabh Pant University of Agriculture & Technology, Pantnagar in July 2014 for Master's degree in Computer Engineering.

Address:

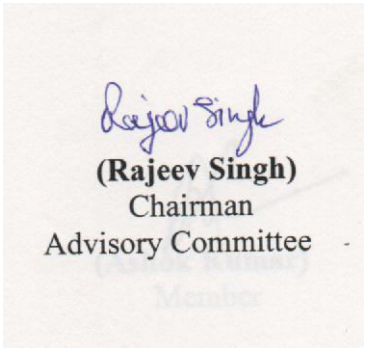
Pradeep Giri
S/o Mr. Hoshiyar Giri
New Nera Baste,
P.O. Siltham - 262501
District - Pithoragarh
Uttarakhand
E-mail ID: pgpradeepgiri@gmail.com
Phone no.: +919690175118

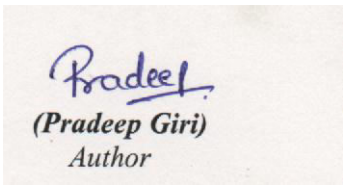
ABSTRACT

Name	**:** Pradeep Giri	**Id. No.**	**:** 48141
Semester & Year of admission	**:** I, 2014-2015	**Degree**	**:** Master of Technology (Computer Engineering)
Major	**:**Computer Engineering	**Department**	**:**Computer Engineering
Thesis Title	**: Human intervened CAPTCHA**		
Advisor	**:** Dr. Rajeev Singh		

CAPTCHA is an abbreviation for Completely Automated Public Turing-Test to tell Computers and Humans Apart. The basic aim of a CAPTCHA system is to prevent Web services from being exploited by hackers. CAPTCHA helps in preventing bots from doing unlawful activities on web pages. The thesis work includes two types of CAPTCHA i.e. Automated CAPTCHA and Human Intervened CAPTCHA (HI-CAPTCHA). The automated CAPTCHA is used for learning the behaviour of bots and for their identification.

In this research work a modified CAPTCHA i.e. HI-CAPTCHA is proposed, which is used for the identification of bot as well as for the identification of the genuine user. All the users inside the classroom are genuine users as they are authorized to use the system. The proposed system is targeted towards enhancing security in mobile based classroom system like attendance system, examination system etc. In this system questions are generated by the admin and the answer of the question varies as per the admin's state and other things that can be identified by the person inside room. Hence bots and invalid users (users outside the class) are not able to detect the answer and system remains secure.





सारांश

नाम : प्रदीप गिरी **परिचायंक संख्या** : ४८१४१

उपाधि : मास्टर ऑफ़ टेक्नोलॉजी (संगणक अभियांत्रिकी)

षठमास एवं प्रवेश वर्ष : प्रथम, २०१४-२०१५

विभाग : संगणक अभियांत्रिकी

मुख्य विषय : संगणक अभियांत्रिकी

शोध ग्रंथ शीर्षक : **मानव हस्तक्षेपित कैप्चा**

सलाहकार : डा० राजीव सिंह

कैप्चा एक पूरी तरह से स्वचलित सार्वजनिक ट्यूरिंग टेस्ट है जो संगणक और मनुष्य के बीच अंतर बताता है। एक कैप्चा प्रणाली का मूल उद्देश्य हैकर्स द्वारा वेब सेवाओं का अनुचित लाभ उठाने से रोकना है। कैप्चा प्रणाली, बॉट्स की वेब पन्नों पर गैरकानूनी गतिविधियों को रोकने मैं मदद करता है। इस शोध कार्य में दो तरह के कैप्चा शामिल हैं जिसमें स्वचलित कैप्चा और मानव हस्तक्षेपित कैप्चा हैं। स्वचलित कैप्चा का प्रयोग, बॉट्स के व्यवहार को सीखने और उनकी पहचान के लिए किया गया है। इस शोध कार्य में संशोधित कैप्चा यानी मानव हस्तक्षेपित कैप्चा को प्रस्तावित किया गया है, जो कि बॉट्स के साथ-साथ एक वास्तविक उपयोगकर्ता की पहचान के लिए प्रयोग किया जाता है।

इस कार्य मैं यह माना गया है कि कक्षा के अंदर सभी उपयोगकर्ता वास्तविक उपयोगकर्ता हैं, जो कि इस प्रणाली का उपयोग करने के लिए अधिकृत हैं। प्रस्तावित प्रणाली का उद्देश्य मोबाइल आधारित कक्षा प्रणाली की सुरक्षा को बढ़ाने की ओर लक्षित है, जैसे की हाज़री प्रणाली, परीक्षा प्रणाली व अन्य। इस प्रणाली में व्यवस्थापक द्वारा सवालों को मोबाइल द्वारा पूछा जाता है। सवालों के उत्तर व्यवस्थापक की स्थिति और कमरे के अंदर उपस्थित व्यक्ति के अनुसार बदलते हैं। इसलिए बॉट और अवैध उपयोगकर्ता (कक्षा के बाहर मौज़ूद उपयोगकर्ता) जवाब पता लगाने में सक्षम नहीं होते हैं और प्रणाली सुरक्षित रहती है।

(राजीव सिंह)
सलाहकार

(प्रदीप गिरी)
लेखक