

การกำหนดประโยชน์ของ CAPTCHA เฝ้าข้อความในประเทศกำลังพัฒนา

นายคุณชิ่ง นอร์บุ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต  
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ  
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์  
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย  
ปีการศึกษา 2555  
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)  
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)  
are the thesis authors' files submitted through the Graduate School.

USEFULNESS DETERMINATION OF TEXT-BASED CAPTCHA IN A GROUP OF  
DEVELOPING COUNTRIES

Mr. Kuenzang Norbu

A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science Program in Computer Science and Information Technology  
Department of Mathematics and Computer Science  
Faculty of Science  
Chulalongkorn University  
Academic Year 2012  
Copyright of Chulalongkorn University



นายคุณชัง นอร์บุ: การกำหนดประโยชน์ของ CAPTCHA

เชิงข้อความในประเทศกำลังพัฒนา : (USEFULNESS DETERMINATION OF TEXT-BASED CAPTCHA IN A GROUP OF DEVELOPING COUNTRIES)

อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ.ดร.ภัทรสินี ภัทร โสศล, 48 หน้า.

CAPTCHA เป็นเทคนิคสำคัญสำหรับการป้องกันระบบจากซอฟต์แวร์ที่ไม่พึงประสงค์ที่ต้องการเข้าถึงระบบคอมพิวเตอร์ผ่านทางอินเทอร์เน็ต อย่างไรก็ตาม เทคนิค CAPTCHA ก่อให้เกิดปัญหามากมายต่อผู้ใช้งาน โดยปกติ ผู้ใช้งานเข้าใช้ระบบผ่านทางคำตอบคำถามหลายชั้น โดยการนำเสนอของคำถามเหล่านั้นมีหลากหลายรูปแบบ รูปแบบต่างๆ ดังกล่าวใช้เวลานาน และยุ่งยาก ดังนั้นการใช้งาน CAPTCHA จึงเป็นประเด็นที่สำคัญ ที่การออกแบบต่างๆ นั้นควรได้รับความสำคัญ การนำเสนอ CAPTCHA ที่ไม่เหมาะสมกับการใช้งานสามารถทำให้ผู้ใช้งานไม่สามารถเข้าใช้งานระบบได้ ในงานวิจัยนี้ได้ศึกษาปัจจัยหลายๆ ด้านเพื่อกำหนดปัจจัยที่มีผลกระทบต่อประสิทธิภาพของการใช้งาน CAPTCHA งานวิจัยนี้ได้ศึกษากลุ่มตัวอย่าง 3 สัญชาติ คือ ภูฏาน อินเดีย และ ไทย งานวิจัยแสดงให้เห็นความแตกต่างบางประการระหว่างกลุ่มตัวอย่าง 3 สัญชาติ ซึ่งส่งผลกระทบต่อความสำเร็จของการใช้งาน CAPTCHA

สาขาวิชา วิทยาการคอมพิวเตอร์และ.....ลายมือชื่อนิติศ.....

เทคโนโลยีสารสนเทศ.....ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....

ปีการศึกษา.....2555

## 5472603023 : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

KEYWORDS: CAPTCHA/TEXT-BASED CAPTCHA

KUENZANG NORBU: USEFULNESS DETERMINATION OF TEXT-BASED CAPTCHA IN A GROUP OF DEVELOPING COUNTRIES. ADVISOR: ASISTANT PROFESSOR .PATTARASINEE BHATTARAKOSOL, Ph.D., 48 pp.

CAPTCHA is an important technique for protecting computer systems from malicious software accessing computer systems through the Internet. CAPTCHA techniques, however, pose many difficulties to users. Users usually gain entry to systems through layers of questions in various presentation styles that are both time-consuming and cumbersome. Therefore, the usability of CAPTCHA is a significant issue that all designers must concern. Unfriendly CAPTCHA presentations can hinder authorized users to gain access to the required system. In this research, many factors are studied to determine factors that affect the effectiveness of using CAPTCHA. The study was performed with three nationalities: Bhutanese, Indian and Thai. This study shows some dissimilarities between these nations that can have impact in the success of using CAPTCHA.

Department: Mathematics and Computer Science Student's Signature.....

Field of Study: Computer Science ..... Advisor's Signature.....  
and Information Technology.....

Academic Year: 2012.....

## ACKNOWLEDGEMENTS

I would like to express my heart felt appreciation to my thesis advisor, Assistant Professor Dr. Pattarasinee Bhattarakosol for her valuable guidance and making her time available for the guidance without which I would not have completed the research. In addition, I would also like to thank the thesis committees, Assistant Professor Dr. Nagul Cooharajanane and Assistant Professor Dr. Ohm Sornil for sparing their precious time for thesis examinations and for their extremely helpful comments. I would also like to thank all reviewers of the AICIT (Advanced Institute of Convergence Information Technology) Conferences for their incredible comments on the published proceedings that brought improvements on this research.

I would like to thank all the members of my family for providing moral supports and helping me to complete my Master Degree. I also would like to thank all Thai friends for providing me help during my course period in the Chulalongkorn University. Moreover I thank all the professors for their wonderful teaching and for sharing valuable knowledge.

Finally, I would like to thank Thai Government especially TICA (Thailand International Development Cooperation Agency) for their meritorious task of providing such wonderful opportunity to study in a prestigious University. Last but not the least I would like thank the NPPF (National Pension and provident Fund) for providing necessary help and the opportunity.

## CONTENTS

	Page
Abstract (Thai).....	iv
Abstract (English).....	v
Acknowledgement.....	vi
Contents.....	vii
List of Tables.....	ix
List of Figures.....	x

<b>Chapter</b>	<b>Page</b>
I. INTRODUCTION .....	1
1.1 Background and Importance .....	2
1.2 Motivation.....	3
1.3 Problem Statement.....	3
1.4 Objectives.....	4
1.5 Scope and limitations .....	5
1.6 Expected Outcomes.....	<a href="#">56</a>
1.7 Thesis Structure.....	6
II. RELATED WORKS AND THEORIES .....	<a href="#">78</a>
2.1 CAPTCHA Introduction .....	<a href="#">78</a>
2.2 Application of CAPTCHA.....	<a href="#">89</a>
2.3 Types of CAPTCHA .....	<a href="#">910</a>
2.3.1 Image-based CAPTCHA .....	<a href="#">1112</a>

2.3.2	Audio-based CAPTCHA .....	13
2.3.3	Video-based CAPTCHA.....	14
III.	METHODOLOGY .....	<a href="#">2021</a>
3.1	Identifying factors .....	<a href="#">2021</a>
3.2	Data Collection.....	<a href="#">2122</a>
3.2.1	Sample Selection.....	<a href="#">2223</a>
3.2.1.1	Bhutan.....	<a href="#">2223</a>
3.2.1.2	India .....	<a href="#">2324</a>
3.2.1.3	Thailand.....	<a href="#">2324</a>
3.2.2	Demographics of Sample.....	<a href="#">2425</a>
3.3	Collecting Technique and Attributes.....	<a href="#">2829</a>
3.4	Data analysis .....	32
IV.	RESULTS.....	33
4.1	General comparisons among countries .....	33
4.2	The Usefulness Analysis .....	<a href="#">3637</a>
4.3	Summarization of the usefulness .....	<a href="#">4744</a>
V.	DISCUSSION AND CONCLUSIONS .....	<a href="#">4846</a>
5.1	Discussion.....	<a href="#">4846</a>
5.2	Conclusions .....	<a href="#">5048</a>
	REFERENCES .....	<a href="#">5250</a>
	APPENDICES .....	<a href="#">5654</a>
	APPENDIX A .....	<a href="#">5755</a>
	APPENDIX B.....	<a href="#">5856</a>

APPENDIX C.....7168

APPENDIX D .....7774

BIOGRAPHY

.....80

77

## LIST OF IMAGES

Figure	Page
Figure	Page .....
Table	Page .....
Table 3.1: Gender-wise classified data	26.....
Table 3.2: Age-wise classified data	26.....
Table 3.3: Occupation-wise classified data	27.....
Table 4.1: Showing Frequency of Internet User	34.....
Table 4.2 Relative comparison between Nationalities on frequency of Internet user	35.....
Table 4.3 Relative comparison between age-group on frequency of Internet user.	35....
The Table 4.3 shows the detailed relative comparison between age-group with regards to	
frequency of Internet User. ....	36.....
Table 4.4 Detailed Relative comparison between age-group on frequency of Internet user	36.....
Table 4.5 Relative comparison between Genders on frequency of Internet user.	37.....
Table 4.6 Relative comparison between occupations on frequency of Internet user	38.....
Table 4.8 Frequency table of CAPTCHA Type	45.....
Table 4.9 Relative percentage of CAPTCHA preferred	45.....
Table B1: Sample collected Details	59.....
Table B2: Indicates the CAPTCHA knowledge of the Samples	59.....
Table B3: The respondents view with regards to CAPTCHA	60.....
Table B4: How the user gets to know CAPTCHA or Source of CAPTCHA	60.....
Table B5: User's selected choice group by nationalities	60.....
Figure 2.1: Samples of symbols based CAPTCHA .....	910
Figure 2.2: Samples of Text- based CAPTCHA.....	1011
Figure 2.3: A Sample of the Image- based CAPTCHA.....	12
Figure 2.4: A sample of the Audio CAPTCHA .....	1314
Figure 2.5: A sample of the Video-based CAPTCHA.....	14

Figure. 3.1 The Internet Users in Bhutan (est 2011)[43] .....	2324
Figure 3.2: Internet Users in Thailand (est 2011) [43]. .....	2425
Figure A1: A kind of CAPTCHA: Circle aptcha .....	57
Figure A2: Example Question based CAPTCHA .....	57
Figure A3: Another Example of Question based CAPTCHA .....	57

## LIST OF TABLES

Table	Page
Table 3.1: Gender-wise classified data	26
Table 3.2: Age-wise classified data	<a href="#">Error! Bookmark not defined.26</a>
Table 3.3: Occupation-wise classified data	27
Table 4.1: Showing Frequency of Internet User	34
Table 4.2 Relative comparison between Nationalities on frequency of Internet user	<a href="#">Error! Bookmark not defined.35</a>
Table 4.3 Relative comparison between age-group on frequency of Internet user.	<a href="#">3435</a>
The Table 4.3 shows the detailed relative comparison between age-group with regards to frequency of Internet User.	
	<a href="#">3536</a>
Table 4.4 Detailed Relative comparison between age-group on frequency of Internet user	<a href="#">Error! Bookmark not defined.36</a>
Table 4.5 Relative comparison between Genders on frequency of Internet user.	<a href="#">3537</a>
Table 4.6 Relative comparison between occupations on frequency of Internet user	<a href="#">Error! Bookmark not defined.38</a>
Table 4.8 Frequency table of CAPTCHA Type	45
Table 4.9 Relative percentage of CAPTCHA preferred	<a href="#">Error! Bookmark not defined.45</a>
Table B1: Sample collected Details	<a href="#">5859</a>
Table B2: Indicates the CAPTCHA knowledge of the Samples	<a href="#">5859</a>

Table B3: The respondents view with regards to CAPTCHA [5960](#)

Table B4: How the user gets to know CAPTCHA or Source of CAPTCHA [5960](#)

Table B5: User's selected choice group by nationalities [5960](#)

# Chapter I

## INTRODUCTION

CAPTCHA stands for “Completely Automated Public Turing Test to Tell Computers and Humans Apart” [13], also known as Human Interaction proof (HIPs). The term “CAPTCHA” was coined in 2000 by Luis von Ahn, Manuel Blum Nicholas J.Hopper, and John Langford. It is a method programmed to identify human and computer bots, by generating a test that only humans can solve easily and computers cannot. The method assumes that any user entering the correct solution is human and those who enters incorrect solution as imitated human.

Basically CAPTCHA is classified into four Categories:

1. Text-based CAPTCHA
2. Image-based CAPTCHA
3. Audio-based CAPTCHA
4. Video-based CAPTCHA

Although, the main purposes of all these types of CAPTCHA are the same as mentioned earlier; however, ~~they are different from each other in implementation it defers from the way it has been implemented.~~ For example, the Text-based CAPTCHA uses alphabets and numbers with some special effects like distortion, adding noise, etc. In the Image-based CAPTCHA, some collective images are clubbed together to form a single image, and ~~then~~ users ~~with will~~ be asked to perform a certain activity ~~accordingly~~. The Video-based CAPTCHA, ~~where~~ ~~in which~~ a video will be displayed and users will be asked to perform a certain task. Lastly, the Audio-based CAPTCHA is designed especially for ~~those~~ groups of people who have vision problems. ~~After An-an Audio-audio will bewas~~ played, ~~and~~ the users ~~will~~ have to identify the sound and ~~will have to~~ type the right input accordingly. More ~~in~~ details ~~of about~~ each type of CAPTCHA are explained in Chapter 2.

Although there are many types of CAPTCHAs as mentioned above, the Text-based CAPTCHA is widely used ~~as~~ because it is easy and comprehensible. The Text-based CAPTCHA has been implemented in many websites in order to prevent intrusion ~~from~~ by automated systems which are trying to illegally gain access to the system.

~~The interesting issue of The~~ the Text-based CAPTCHAs ~~are~~ is the very useful ~~for~~ protection of intrusions from automated bots. The usefulness of CAPTCHA depends on the design of CAPTCHA ~~and the design of CAPTCHA depends~~ which bases on design features and other factors like physical characteristics, mental capability of the user.

The above mentioned factors may ~~differ from individual to vary in~~ individual ~~differences, from one location to another location, and from one country to another country and~~ different locations and countries. To evaluate the effect of ~~the~~ above factors, three countries Bhutan, India, and Thailand, which have varying influence of factors in each country, were selected for the study ~~that have varying influence of factors in each country~~. Based on the factors mentioned ~~about~~ above, the usefulness of the Text-based CAPTCHA has been determined and the results are detailed in Chapter 4.

## 1.1 Background and Importance

Currently, in this modern world, computers have spread out over the globe and ~~its~~ their utility ~~utilities~~ has ~~have been~~ grown to an extent in the world where the minuscule malfunction of the computer system can lead to huge financial losses. With the invention of the Internet, the usage of computers has been become inevitable. The world's information flow has widely changed with the implementation of the Internet. As a consequence, the Internet becomes a significant resource for the information. Moreover, the use of the Internet accelerates the business process competition in a global presently. In order to gain their successes in the same business, some business rivalries have propagated malicious computer programs to disrupt their rival's business processes. Additionally, various malwares were distributed to gain unauthorized access to disclosed company information.

Besides, in 1999, Slashdot [31] published a poll asking visitors to choose the graduate school, which had the best program in computer science? To gain the fame, the students from two universities, Carnegie Mellon and MIT, created automated programs called bots to vote repeatedly for their respective schools. While these two schools received thousands of votes, other schools had only a few hundred each. These acts have necessitated the creation of the system to be differentiate between human and malware inputs. Consequently, in the year 2000, researchers introduced a method, called Completely Automated Public Turing Test, to Tell Computers and Human Apart, which is-was commonly known as CAPTCHA [1], in order to prevent this kind of intrusion from unwanted programs. Afterward, free email service providers like Gmail and ~~yahoo~~-Yahoo mail implemented the CAPTCHA to prevent automated mail registrations and many other webmail providers followed them.

## 1.2 Motivation

As mentioned previously that the Internet has spread all over the world. In developing countries, technologies play an important role in development and economic growth. Developing countries like Bhutan where the Internet was introduced in the late 1999 which could be one of the last countries to introduce the Internet. However, the growth rate in technology and people-usingthe usage of the Internet in the developing countries has been marked remarkable. With this remarkable growth in the technology usages, the security has become a great concern. One of the major security concerns is CAPTCHA as many bots or unwanted programs try to access the system illegally. Thus, it is interesting and necessary to conduct study on the usefulness of CAPTCHA, particularly, in some of the developing countries like Bhutan, India and Thailand. This research focuses more in-on the Text-based CAPTCHA as it is the most commonly applied to the websites. Therefore, the study would reveal determination of usefulness of the Text-based CAPTCHA in some of the developing countries.

## 1.3 Problem Statement

Presently, CAPTCHA has been used in many websites for the purpose of protecting its-theirs information from intruders or bots or mimicking human users. Therefore,

CAPTCHA method plays an important role in shielding system against malwares or bots (automated scripts) and prohibited users. However, occasionally even the legal users ~~are do not also not granted~~ access to the system due to the unfavorable CAPTCHA design. ~~This-The~~ unfriendly nature of the CAPTCHA may degrade the usefulness of it. Therefore, ~~it-this~~ is a concern issue to consider usefulness when designing the CAPTCHA.

#### 1.4 Objectives

CAPTCHA, is a technique developed to filter non-human users from real human without using their credentials. It has been seen that there are many usefulness issues with every kind of software or hardware developments. CAPTCHA also faces the same usefulness issue. Usefulness may vary from place to place and from individual to individual. Factors have been broadly categorized into human physical characteristics, mental factors and the design features. These factors have differing ranges of effects in the usefulness of the CAPTCHA. Further, the various types of CAPTCHA schemes may exhibit some differences and some common factors.

~~However~~~~Therefore~~, this study will focus on determining the usefulness of the Text-based CAPTCHA in developing countries. As discuss earlier, factors such as human characteristics ~~which~~ differ from individual to individual and more importantly from country to country. Examples of ~~such~~ human characteristic factors are ~~such as~~ blind or visually impaired users who may not be able to use Text-based CAPTCHA. Thus, in such case the Text-based CAPTCHA may not be of much use, so the Text-based CAPTCHA is unfeasible. Likewise there are other mental factors such as experiences and knowledge of users with regards the use of the CAPTCHA.

Besides, the CAPTCHA design factors seem to have a veiled impact on the success of the CAPTCHA usage. The design factors of the Text-based CAPTCHA are such as characters' colors, background colors, fonts, and sizes, special effects like distortion and noises. These design factors are all contributing factors towards the usefulness of the Text-based CAPTCHA. Nevertheless, every individual has different perception about the usefulness of CAPTCHA. Therefore, there are many influencers that affect the usefulness of the Text-based

CAPTCHA. So, the design factors and other factors of usefulness should be taken care to achieve the satisfactory usability levels of the Text-based CAPTCHA.

Accordingly, this study aims to clarify the usefulness of the Text-based CAPTCHA in the underdeveloped countries in account of the factors mentioned above. The selected countries in this study are Bhutan, India, and Thailand. These countries have different levels of the technology maturity. It is imperative that computer-users of these developing countries may have different levels of exposure towards the Internet. Therefore, factors that affect the usefulness of the CAPTCHA may also vary across the variation of countries.

## **1.5 Scope and limitations**

CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Human Apart), a human user identifier from bots and other computer generated activities. There are many different types of CAPTCHAs being implemented over time but the Text-based CAPTCHA has been one which is widely used. The study is based on the survey-based questionnaire over three countries. The following list is the scope and limitations of this study:

1. The determination the usefulness of the Text-based CAPTCHA only.
2. The work focuses on simple and less distorted of the Text-based CAPTCHA.
3. The sample groups in this evaluation are limited to three nationalities (Bhutanese, Thai and Indian).
4. The sample groups are between 10 to the age of 60 years and above.
5. The study does not cover vision impaired samples.

## **1.6 Expected Outcomes**

The study of the usefulness of the Text-based CAPTCHA conducted using questionnaire survey. The study is limited to the respondents from three countries namely Bhutan, India and Thailand. The outcomes of this study were aimed at the following points:

1. To determine factors affecting to the usefulness of the Text-based CAPTCHA on the participants from Bhutan and comparing with other two nationalities (Thailand and India) which are technologically advanced than Bhutan.
2. To determine the variation in the usage of CAPTCHA and the cause.
3. To know the users' view regarding to the use of different CAPTCHA designs for different age groups.
4. To determine the CAPTCHA scheme is popular among all three nationalities (Bhutanese, Thai and Indian)

## 1.7 Thesis Structure

The structure of the thesis is divided into chapters and the each chapter contains as follows.

- **Chapter 1:** The introduction for this study ~~that~~ includes the problem statements, background and motivations. The objectives of this study and details of study conditions are stated.
- **Chapter 2:** The related works and theories that are further broken into introductory, usage, type and weakness and strength and history of CAPTCHAs, all are presented.
- **Chapter 3:** This chapter basically describes the methodology like the research method and details.
- **Chapter 4:** This chapter contains results of the data capturing and analysis
- **Chapter 5:** This contains the discussion and the conclusion

# Chapter II

## RELATED WORKS AND THEORIES

In this chapter, the basic knowledge and reviews of some related works are presented. The basic knowledge is such as the definitions, types, and usage of CAPTCHA. Then, the literature reviews related to the research topic are elaborated. Details of such contents are described below.

### 2.1 CAPTCHA Introduction

The computer technologies and applications development leads to the improvement of the ~~Internet~~internet technology, a technology with worldwide accessibility. Every country around the world uses the ~~Internet~~internet applications for their daily lives and activities, such as communication media, business processes, and educations. The Internet has advantages that support daily activities as mentioned earlier easy and faster. However, there is a number of issues in using the Internet.

One of such issues is the security issue. Over years, many security measures were invented and implemented either in the forms of hardware or software based. The security was implemented in order to protect information from illegal users and bots. Some examples of such security methods are biometric security, firewall security, and antivirus.

Sometimes later with the improvement of bots' deciphering methods, activities over the Internet became confusing to differentiate because the bots' activities are similar to that of human. Thus, the security technique in recognizing a user and confirming that the user is human has become so essential. In pursuit of such validation of genuine human-to-human communication, the theoretical method was founded by Moni Noar (1996) [32].

The term CAPTCHA was invented in the year 2000 [13] by Luis Von Ahn from Carnegie Mellon University. It was implemented to control or prevent the malicious software that is designed to attempt unauthorized accesses to computer resources and to validate human-user. It was yahoo to implement the CAPTCHA for the first time and those CAPTCHA

were basically Text-based. The most common Text-based CAPTCHA used are EZ-Gimpy, Gimpy, Baffle-text, Pessimial-Print and Scatter Type [30]. Besides the Text-based CAPTCHA, there are other kinds of CAPTCHA, namely Image-based, Audio-based and Video-based.

## **2.2 Usage of CAPTCHA**

CAPTCHA is a program that produces and executes tests which are designed to be tough for malwares to pass through but it is easy for humans. The main purpose of CAPTCHA is to ensure that requested services or resources are generated for human and not for any kinds of automated computer programs. Thus, it attempts to identify legal users from illegal or virtual users. Accordingly CAPTCHA has been used as a front line web security where humanly activities need to be verified. Examples of these are registration forms, online polling [1], and even e-banking [2]. In addition, it has been used to prevent spammers [3], Spyware [4], Denial of Service attacks [5], and also used to identify the child users [6]. CAPTCHAs are also used to prevent dictionary attacks in password systems and to block search engine bots.

The use of CAPTCHA can be experienced when a user creates new webmail accounts in Gmail account, Hotmail and Yahoo mail. Additionally, it is also used in blogs, forums, Wikis and every e-commerce website. Furthermore, it has been observed that CAPTCHA is used in Facebook as Heuristic checks to restrict the overused of its services beyond some threshold. Not only does CAPTCHAs help preventing abuse, it also helps in preserving knowledge by digitizing some old books. The best and live example of such contribution is through the reCAPTCHA project. There has been an attempt to digitize contents of old books using OCR software. However, some certain words are not OCR-readable because of typical printing styles with faded ink and yellow pages. For that reason, Von Ahn and coworkers [31] have used those words to display in CAPTCHA so that humans could help interpreting the words. This method is named as reCAPTCHA. This reCAPTCHA improved the performance of digitizing old printed contents and making that knowledge more reachable to the public. In fact, the used of those words which were not properly identified by the OCR system in CAPTCHA and making use of the ability of human being able to read those words has contributed lots towards the feeding of word in digitized pool of words. Thus, CAPTCHA can

be applied for many purposes besides front line security. Consequently, CAPTCHA is very useful in this modern world.

## 2.3 Types of CAPTCHA

CAPTCHA is classified in many styles. Some categorize CAPTCHA based on its presentation forms or appearance while some groups of CAPTCHA are based on the technology. The various types of CAPTCHA are OCR-based (Optical character recognition) and Visual Non-OCR-based types. These are technology-based classification. Other types of CAPTCHA are Non visual [7], Text-based CAPTCHA [8], Audio-based CAPTCHA [9], Image-based CAPTCHA, Puzzle-based CAPTCHA [10], QRBGGS (Quantum Random Bit Generator Service) [11], Cartoon-based CAPTCHA [16] and Question-Based CAPTCHA [12] which basically can be widely classified by their appearances. Moreover, new type of CAPTCHA is also evolving such as symbol-based CAPTCHA as shown in the Figure 2.1 and as well in Appendix A.

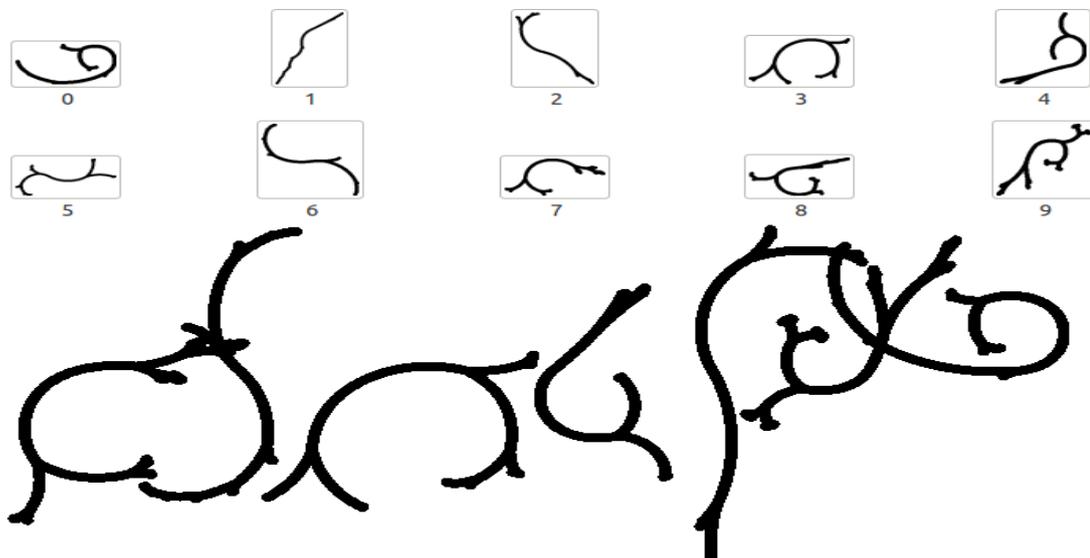


Figure 2.1: Samples of symbol-based CAPTCHA

### 2.3.1 Text-based CAPTCHA

The Text-based CAPTCHA is designed to provide tests where users are confronted with texts or digits with distortions; the user needs to identify and input the same value to gain access to intended resources. Some examples are Gimpy, EZ-Gimpy, PessimialPrint, Baffle Text and Scatter Type. Gimpy [20] was the first type of CAPTCHA that Yahoo has implemented. This type of CAPTCHA is implemented by the webmail service providers like Google and Hotmail to allow the real human being to register for the e-mail services. Moreover, this type of CAPTCHA is more commonly implemented than any other types. EZ-Gimpy is another distorted Text-based CAPTCHA where users have to identify multiple words [20].

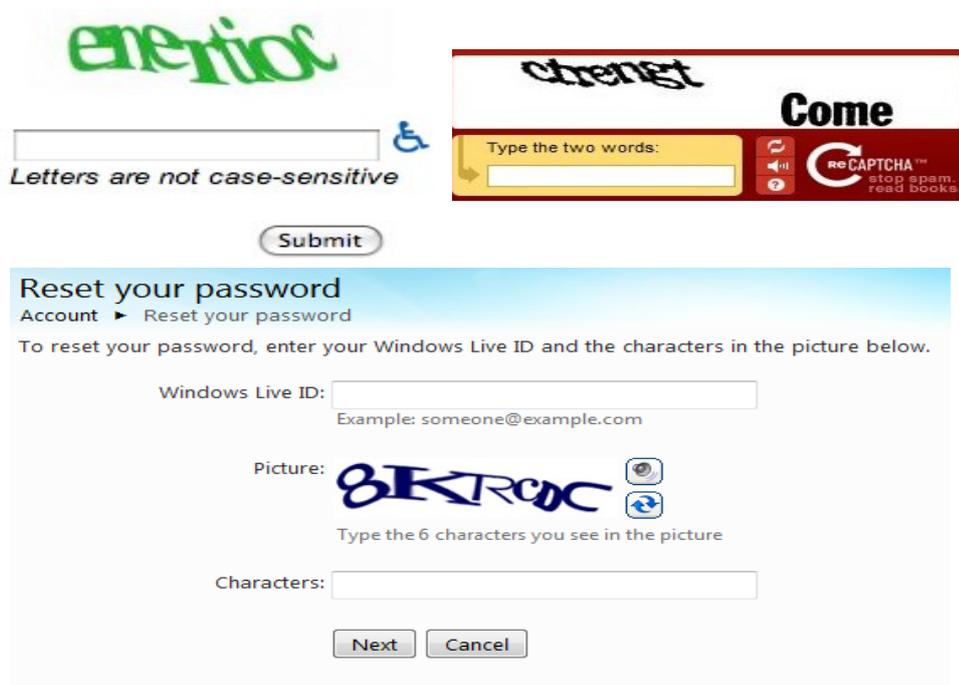


Figure 2.2: Samples of Text- based CAPTCHA

Another example of CAPTCHA is the Logical Questions Text-based CAPTCHA, an example of such kind is 8 plus 2 is what? The user inputs the answers to these questions in the text-box provided for the input purpose. Some samples of the Text-based CAPTCHA are shown in Figure2.3 and Figure 2.4.

## Eyestrain Captcha



Figure 2.3: Eyestrain Text- based CAPTCHA



### 2.3.2 Image-based CAPTCHA

This Image-based CAPTCHA displays an image, the user has to answer the question based on the presented image to accomplish the test. Currently, there are various techniques of the Image-based CAPTCHA such as the confident CAPTCHA, Bongo CAPTCHA, and ESP-PIX, etc.

The confident CAPTCHA is the same kind of the Image-based CAPTCHA; it asks visitors to click and identify specific pictures in order to filter spams and bots passing through websites.



Figure 2.5: A Sample of the Image-based CAPTCHA

The Bongo CAPTCHA is another example of the Image-based CAPTCHA where users are asked to solve a visual pattern recognition problem [21]. It uses two sets of images and each set contains some specific characteristics; one set might be boldface while another is differed. The user must be presented with an image and its belonging set must be identified by the user.

Unlike from the above two techniques, the ESP-PIX shows four images to the user and asks the user to choose a word related with all images from the dropdown list. Similar to the ESP-PIX, the ASIRRA (Animal Species Image Recognition for Restricting Access: similar to KittenAuth as shown in Figure 2.3) requires the user to identify all the pictures of cats from a set of 12 images containing cats and dogs.

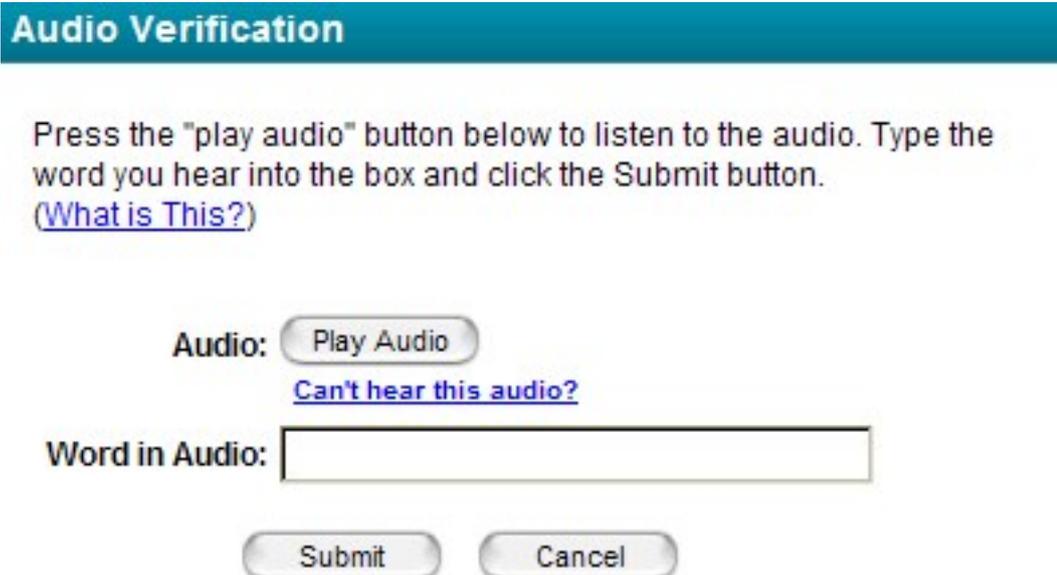
However, there is another Image-based technique that users have to draw an outline around a specific object in the image. This technique is named as SQ-Pix.

Some more examples of the Image-based CAPTCHA are Picatcha, KittenAuth, reCAPTCHA and IMAGINATION. However, due to lack of huge image database, the Image-based CAPTCHAs are considered insecure. Hence, the Image-based CAPTCHA is not that popular as the Text-based CAPTCHA.

### 2.3.3 Audio-based CAPTCHA

The Audio-based CAPTCHA is one of the CAPTCHA types. This technique asks users to type a statement or words that have been pronounced by the program. The sample of the Audio-based CAPTCHA is shown in the Figure 2.4. This technique is developed to support users with the poor vision problem. Thus, this technique is often combined in the same interface as a visual CAPTCHA. One sample of this Audio-based CAPTCHA is the reCAPTCHA system.

Audio-based CAPTCHA are much difficult than other techniques because the original sound is always contaminated with other background sound. Thus, users usually cannot pass the test. Moreover, the speech of the testing audio is too fast to remember so users cannot key the text correctly. Consequently, it is not that commonly used as the Text-based CAPTCHA.



**Audio Verification**

Press the "play audio" button below to listen to the audio. Type the word you hear into the box and click the Submit button.  
([What is This?](#))

Audio:  [Can't hear this audio?](#)

Word in Audio:

Figure 2.6: A sample of the Audio CAPTCHA

### 2.3.4 Video-based CAPTCHA

CAPTCHA security is constantly changing with the growth in the ability of the hackers. So the Video-based CAPTCHA is proposed to protect the system from hackers. This



Figure 2.7: A sample of the Video-based CAPTCHA

technique uses animations and moving backgrounds to block the automated identification of CAPTCHA by malwares.

It has come into existence lately; one example is NuCaptcha which is shown in Figure 2.5. Its videos run in a small-embedded screen. It comes in a variety of themes and shows a predetermined, constant white text (“Security” followed by a short string of red letters, “8HE”). The user is asked to type in a box below the screen, as displayed in the figure. According to the example in Figure 2.5, the CAPTCHA is presented in a simple form without noises except the movement of CAPTCHA. Thus, this technique is much simple and easy for users to identify CAPTCHA while the automated software is unable to do the task. Therefore, it is also said to be more secure than other techniques mentioned previously.

### 2.3.5 Modifications of CAPTCHA

Instead of applying a simple CAPTCHA as mentioned above, there is a new proposed method that combines the simple CAPTCHA with another factor. For example, a four-panel cartoon CAPTCHA is a modified Image-based CAPTCHA with human recognition system. The checking mechanism of this new CAPTCHA is to display four panels cartoon in a disorder mode and asks users to re-arrange the presented panels in the right order. Thus, the human cognitive power can be applied to differentiate between malware and the legitimate human.

Another method is the combination between the Tex-based CAPTCHA with the keystroke mechanism. This method captures the typing time when the Text-based CAPTCHA is displayed and typed by the user. The timestamps can be used to differentiate between human and malwares. This method is called bio-detection functionalities [17]. Likewise, SEMAGE (Semantically Matching Image) a novel Image-based two-factor CAPTCHA system that is built upon the idea of semantic relationship between images [29]. Further, CAPTCHA using Jigsaw Puzzle [10] that implements the jigsaw puzzle logic where only two pieces of images are misplaced from their original place. Likewise there are many others types.

## 2.4 Related works

Although, the Text-based CAPTCHA has been used widely due to its easiness, it encounters many issues related to robustness and usability. Since the Gimpy was the first Text-based CAPTCHA, it was broken by the implementation of the image processing technology where the presented text was segmented and recognized using the image processing technology [14] [15]. Using this methodology, Jitendra Malik and Greg Mori were able to break the E-Z Gimpy with a 92% success rate [20]; similarly, Yan and El Ahmad had a success rate of 61% in breaking the MSN Scheme [25].

Despite having taken care of all the weakness and factors that have contributed towards decrypting the CAPTCHA, the CAPTCHA is still solved easily through the method called the 3<sup>rd</sup> party human attack. It has been found that more and more such attacks are being carried out. In fact, in India a business is conducted by employing some people as human CAPTCHA solvers at \$2 for every thousand CAPTCHAs solved [27]. To counter such practices another type of CAPTCHA was introduced called iCAPTCHA [24] where this CAPTCHA works interactively taking account of the timeout values for solving it.

According to the distortion technique in the Text-based CAPTCHA, it leads to the usability problem of users [28]. The distortion creates such confusing characters and some of the examples such misinterpretation are stated below:

- Distortion of letter and digit to some extents resulted in the following confusion among the letters and digit:
  - 0 from o
  - 6 from G and b
  - 5 from S/s
  - 2 from Z/z
  - 1 from I
  - 5 from 6 or 7
  - 7 form I
  - 8 from 6 or 9.
  
- Distortion of letters under some angle has resulted in the following confusions:
  - “vv”(double v without space) can look like “w”
  - ”cl” can look like “d”
  - ”nn” can look like “w”
  - ”rn” can look like “m”
  - ”rm” can look like “nn”
  - ”cm” can look like “an”

— Distortion of Characters and letter have resulted in the following difficulties:

- The difficulties of being not able to differentiate an arc from character such as “J”, “7” and “L”.

It is also notice that contents of the Text-based CAPTCHA have some effect on usability problem. Occasionally, it happened that when the contents are generated randomly, it formed some offensive words such as “Negro”. Likewise the word “Wait” in the contents made first time user to wait for hours and similarly the word “Restart” made the user restart the computer that the user was using.

Additionally, Presentation of the CAPTCHA was also found to have some impact on the usability problem. The presentation includes the font type, the size used, and the font color used. The researcher mentioned that when font colors are used properly can enhance user interface design but if not used properly, it can hamper efficiency the usability.

Sam Hocevar discovered some factors that help in accelerating the breaking of the CAPTCHA scheme [26]; those factors are constant font, constant character position, non-textured background, no distortion, no perturbation, constant rotation and weak color. As such, the CAPTCHA’s designing factors such as its invariant weakness, native language, addition of noise and distortion and effect of text colors in CAPTCHA are of much concern [16][17].

Yan and El Almad [28] confirmed that those with no background in Latin alphabets had more serious problems in interpreting CAPTCHA than those with the background. However, this work also found that CAPTCHA text’s color does not help in strengthening the security issue of CAPTCHA. In fact it might have a negative effect on usability when the background color and the text font color are the same. Efforts to improve the security of the CAPTCHA by adding noise and distortion have brought in negative impact on usability by reducing human readability of the CAPTCHA. Thus, securing with such add-on reduces human’s ability to solve CAPTCHA resulting in barring the legitimate users from resources.

The inflexible impacts from CAPTCHA can be in seen in [41] where world’s largest online ticket retailer, Ticketmaster, has already decided to stop using CAPTCHA and

started using another software created by New York start-up Solve Media. This software provides similar service as CAPTCHA that asks for well-known phrases, or simple multiple choice questions. Kip Levin, Ticketmaster's executive vice president of eCommerce has commented that the average time to solve a CAPTCHA puzzle was 14 seconds while the new system spends only an average of seven seconds to figure out. Thus, the CAPTCHA designer must realize that the average time to solve a CAPTCHA is a very important factor and should be taken care of during the design phase of CAPTCHA in order to avoid losing users like Ticketmaster. The study has shown the impact of design factors and its consequences related to the usability of the Text-based CAPTCHA.

Similarly, Chen-Chi [18] conducted comparative usability findings among different types of CAPTCHAs, based on the performance metrics which include the finishing time, the rate of typing errors, the rate of timeout, the rate of giving up, and the rate of repeat typing. The finding was that the usability of different CAPTCHA Schemes remains relatively consistent. However, the result from the comparative study performed between the Image-based CAPTCHA and the Text-based CAPTCHA shows that the Image-based CAPTCHA's (Cortcha) has an overall accuracy rate of 86.2% with 18.3s average CAPTCHA solving time while Google's Text-based CAPTCHA has 82.8% accuracy rate with 7.9s CAPTCHA solving time [19]. This demonstrates that the usability of CAPTCHA depends on CAPTCHA Schemes. As it is known that different CAPTCHA Schemes have different variant levels of noise such as interferences of lines, arch, dots, and distortion. Thus, the study indicates that the usability scale depends on types of the CAPTCHA schemes.

The concept of usefulness plays an important role in designing a CAPTCHA which is easy for human to solve but hard for bots. According to the book of Jakob Nielsen [22], usefulness is the question of whether the system can be used to achieve some desired goal. This can further be broken down into the two categories of utility and usability.

The utility is the question of whether the functionality of the system in principle can do what is required to do while the usability is the question of how well users can use that

functionality. Moreover, the usability is not a single, one-dimensional property of a user interface.

Usability has multiple components and is traditionally associated with these five usability attributes: learnability, efficiency, memorability, errors and subjective satisfaction. There are several methods to measure the usability feature, such as computer system usability questionnaire (CSUQ) and the NASA Task Load Index (NASA-TLX). These two methods can measure the performance metrics, such as task completion time, number of errors made by participants, eye gaze measures (numbers of fixations and fixation duration) and subjective measures (satisfaction) [23].

However, Christos [36] considered the efficiency by recording numbers of tries needed to solve a CAPTCHA and perceived difficulties by finding out the obstacles faced when solving a CAPTCHA.

Likewise, Kapi [23] used factors such as efficiency by measuring the average time taken to complete the task, recording the numbers of errors made, measuring the average eye fixation duration for perceived difficulties, and satisfaction using computer system usability questionnaire (CSUQ). This CSUQ is broken into system usefulness, Information Quality, and Interface Quality. Kapi used NASA-TLX (NASA Task load index) that was used to measure the workload, including mental demand, physical demand, and frustration; while Yan's method considered efficiency, error and satisfaction by measuring the accuracy, response time and the subjective satisfaction [28].

However, when the usefulness is concerned with CAPTCHA, there are two categories of factors that have maximum effect on the use of CAPTCHA. These factors are robustness and usability. Although there are various types of CAPTCHA, this study will focus on the usefulness of the Text-based CAPTCHA so that different factors can be discovered.

Next chapter will elaborate the methodology that is applied for recovering the usefulness of the Text-based CAPTCHA.

# Chapter III

## METHODOLOGY

The Text-based CAPTCHA was chosen for this study, as it is the only CAPTCHA that is most commonly used over the websites. Moreover, it is used in Gmail registration which is used by many e-mail users. The study is based on opinion of the respondents from three different countries (Bhutan, India and Thailand) which have variant Information Technology development state. Therefore, the views and opinion were collected through means of the online survey questionnaire. The details of methods followed in determining the usefulness of the Text-based CAPTCHA are as follows.

### 3.1 Identifying factors

From the study of other research papers, they found that CAPTCHA has relatively good learnability and memorability. So these two factors are not taken into account for this study. However, the factors that are taken into account for this study are broadly categorized into three groups according to users as follows.

- 1) experience and knowledge
- 2) perceive difficulties
- 3) satisfaction

These three factors are considered because of the technology advancement. Since people will have opportunity to experience in the technology usage for their better lives, therefore, this experience can create the attitude towards the technology. Moreover, people also have different knowledge; thus, the technological skill will be different. So, both factors, experience and knowledge, can affect to the users' appreciations that can reflect to the usefulness of CAPTCHA. Furthermore, the knowledge of users can separate the users' perceptions that refer to the ability in classifying or determining the presented characters of the Text-based

CAPTCHA. In addition, the attitudes gained from using technology with diverse skills also lead to the varieties of appreciations or satisfaction towards CAPTCHA.

### **3.2 Data Collection**

The Respondents were selected from the three countries namely Bhutan, India, and Thailand. The aim of the survey is to find out factors determining to the usefulness of the Text-based CAPTCHA. The chosen data collection tool is self-managed questionnaire and made available every time and anywhere. This type of collecting data does not require personal contacts. Respondents are asked to fill in the questionnaire by reading the instructions on their own. The questionnaires were distributed with two different options. In the first option, questions were sent by an e-mail with a link to an online survey questionnaire. In the second option, they were distributed in the social media like Facebook amongst the groups; friends were asked to pass on the questionnaire with other common friends. The data were collected through the online method in the span of two months.

There are two reasons that justify the use of those two methods. The advantage of the online survey is that it's low costs and the ability to cover a large group of respondents in a short period of time. The response rate was low when survey links were sent by email that is due to the lack of control and interaction with the respondents. The response period is not definite, thus, the response may be procrastinated and at the end respondents might even forget it. Besides that, some respondents, who are not frequent Internet users might not open the mail account for a long time, by time the respondents open the mail it might be mixed up with other mails and might overlooked the mail containing survey link. Even if the respondents see it and submit the survey, it might run out of the survey stipulated period and their responses become useless. In order to increase the response rate of the survey, questionnaires that were distributed to friends over the Facebook were kept reminding repeatedly until the survey questionnaire is responded. This method helped in collecting data from respondents who are fully forgotten in

submitting the survey. Furthermore, all close friends in 'Facebook' were contacted personally over chat and were asked to complete the survey immediately. Similarly, other acquaintances were requested for the early responses.

Since the use of social media is applied, the data collection is much easier and faster. Moreover, the duplicity must be avoided so friends and acquaintances were requested not to submit the survey in the case that they had come across the same survey through common friends.

The distributed data collection was processed to collect information on the following related facts: demography, information related to CAPTCHA like knowledge and experiences of the respondents on CAPTCHA, difficulties in using CAPTCHA, CAPTCHA style and colors, types of CAPTCHA, and views on the usefulness of the CAPTCHA.

### **3.2.1 Sample Selection**

It is very important to select samples that can distinguish the differences in using CAPTCHA based three factors mentioned in the previous paragraph. Therefore, different countries which have completely different characteristics are selected to determine the usefulness of the Text-based CAPTCHA. Three countries that are Bhutan, India and Thailand have been selected for this experiment as there are significant differences in their characteristic as described below.

#### **3.2.1.1 Bhutan**

Bhutan is a small landlocked country, sandwich between India and China. It is technologically less developed country. Bhutan's total population is around 716,896 (2012 Est.)[35]. Bhutan has literacy rate of 47% [40] which consist 60% of Male and 34% of Female. Bhutan has launched Internet in 1999. Bhutan had 500 Internet users in 2000, and in 2011, the users have reached to 150,548 users and the Internet penetration rate is estimated at 21%.

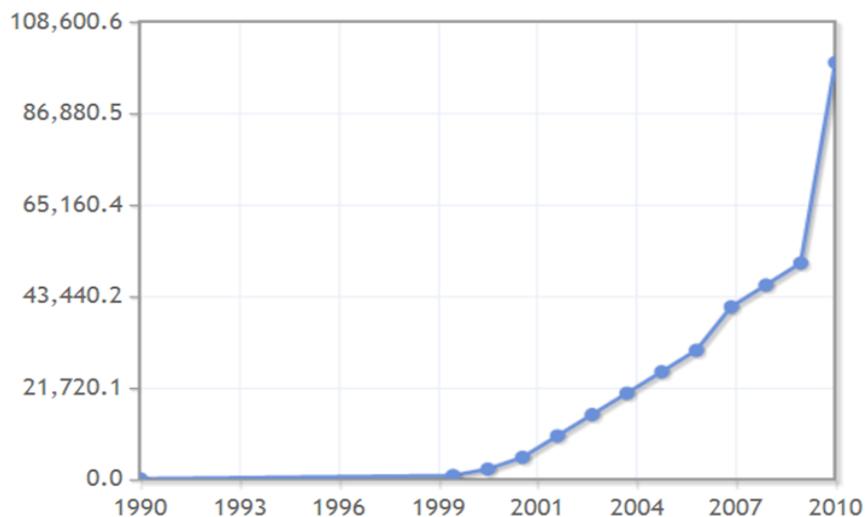


Figure. 3.1 The Internet Users in Bhutan (est 2011)[43]

### 3.2.1.2 India

Indian is one of the large countries in the South East Asia with population of 1 Billion (1,205,073,612) [35]. It is more technologically advanced than Bhutan. India launched Internet in 1995 with 5 million (5,000,000) Internet users in 2000 and the Internet users have reached to 137 million (137,000,000) in 2011. However, the penetration rate of 11.4% which is lower than Bhutan. India has the literacy rate of 74.04% which includes 82.14% males and 65.46% of Females.

### 3.2.1.3 Thailand

Thailand is technologically more developed than India and Bhutan. It covers 513,115 Km<sup>2</sup>, and has population around 67 million (67,091,089) in 2012. However, Thailand experiences the Internet in the year 1996, a year later than India. Thailand had 2 million (2,300,000) Internet users in 2000 and has grown to 20 million (20,100,000) Internet users as of Dec.31, 2011. It is observed that Thailand has high Internet penetration rate at 30.0% when

compared with Bhutan and India. Moreover, the literacy rate is 92.6% out of which 94.9% males and 90.5% females.



Figure 3.2: Internet Users in Thailand (est 2011) [43].

### 3.2.2 Demographics of Sample

316 samples from three nationalities were collected and stored in the database out of which were 181 Bhutanese, 76 Thai, and 59 India. However, 36 respondents were classified as the defect samples during the data cleaning process. Therefore, only 280 respondents remained for this study.

However, the sample collection was performed from three countries that have significant differences in the sample numbers. Thus, these different sizes of samples may lead to the inaccurate conclusions. Therefore, the random process by SPSS was performed in which 48 samples each were selected from the pools of Bhutanese and ~~Thai samples~~ Thai sample to match the ~~Indian samples~~ Indian sample as to have better result of analysis. Consequently, the number of samples for the study is 48 samples from each Nationality. The sample distributions of the selected samples were found as in the following paragraphs.

The sample distributions of Nationalities and genders are depicted in Table 3.1. The sample contains a little bit more men than women. This sample contains respondents in the age range of 10-60 years and the distribution of age groups is depicted in Table 3.2. The majority of respondents is within the age group of 26-40. Furthermore, the respondents are from different ranges of occupation as indicated in Table 3.3.

Table 3.1: Age wise classification of Samples

Age	Nationality		
	Bhutanese	Indian	Thai
10-25	9 (19%, 6%)	15(31%, 10%)	31(65%, 22%)
26-40	35(73%, 24%)	26(54%,18%	17(35%, 12%)
41-60	4(8%, 3%)	7(15%, 5%)	0
	48	48	48

Referring to Table 3.1, the samples contain respondents in the age range of 10-60. It shows that there are 144 respondents which are constituted by 48 samples each from Bhutan, India and Thailand. According to the age-group, there are 9 Bhutanese in the age-group of 10-25, which is 19% of the ~~Bhutanese samples~~Bhutanese sample and 6% of the entire samples. 35 respondents are in the age-group of 26-40 (73% of the ~~Bhutanese samples~~Bhutanese sample and 24% of entire samples). 4 respondents are in the age-group of 41-60, which involve 8% of the ~~Bhutanese samples~~Bhutanese sample and 3% of the sample population. However, there is no sample in the age-group of 61 and above.

The ~~Indian samples~~Indian sample are made up of 15 respondents are in the age-group of 10-25 (31% of ~~Indian samples~~Indian sample, 10% of entire samples), 26 respondents are in the age-group of 26-40 (54% of ~~Indian samples~~Indian sample, 18% of entire samples), 7 are in the age-group of 41-60 (15% of the ~~Indian samples~~Indian sample, 5% of the entire samples), and none is in the age group of 61 and above.

When the Thai sample is considered, 31 respondents are in the age-group of 10-25, consisting of 65% of the ~~Thai samples~~Thai sample and 22% of the entire samples, 17 samples are in the age-group of 26-40, constituting 35% of the ~~Thai samples~~Thai sample and 12% of the entire samples population. However, there is no sample in the age group of 41-60 and 61 and above.

Table 3.2: Gender-wise classified data.

Gender	Nationality			Total
	Bhutanese	Indian	Thai	
Male	26(54%, 18%)	20(42%, 14%)	23(48%,16%)	69(48%)
Female	22(46%, 15%)	28(58%, 19%)	25(52%, 17%)	75(52%)
Total	48	48	48	144(100%)

Referring to Table 3.2 the gender consists of 69 males and 75 females. Moreover, The majority of respondents are within the age group of 26-40. Furthermore, the respondents are from different ranges of occupation as indicated in Table 3.3. Referring to Table 3.1, there are 26 Bhutanese males which are counted as 54% of ~~Bhutanese samples~~Bhutanese sample, 38% of the total males in this survey and 18% of the entire samples. Bhutanese sample contains 22 females which is 46% of Bhutanese sample and 15% of the entire sample and 29% of total female sample.

Similarly, Indian sample consists of 20 males (42% of the Indian sample and 14% of the entire sample and 33% of total male sample). The sample contains 28 females constituting 58% of the Indian sample and 19% of the entire sample population and 37% of the total female samples.

Likewise, ~~Thai samples~~Thai sample comprises of 23 males (48% of the ~~Thai samples~~Thai sample, 16% of the entire samples and 23% of the total male sample). The rest 52% of the ~~Thai samples~~Thai sample are females which count to 25 females (17% of the entire samples and 33% of the total female population).

Table 3.3: Occupation-wise classified data

Occupation	Nationality			Total
	Bhutanese	Indian	Thai	
corporation/private	7(15%, 5%)	5(10%, 4%)	5(10%, 4%)	17(12%)
Government	19(49%, 13%)	5(10%, 4%)	2(4%, 1%)	26(18%)
Business	3(6%, 2%)	8(17%, 6%)	3(6%, 2%)	14(10%)
Information Technology	5(10%, 4%)	4(8%, 3%)	11(23%, 8%)	20(14%)
Others	3(6%, 2%)	0	2(4%, 1%)	5(4%)
Student	11(23%, 8%)	26(54%, 18%)	25(52%, 17%)	62(43%)
Total	48	48	48	144(100%)

Referring to Table 3.3, from the table, the respondents are distributed in various occupation groups. There are 7 Bhutanese working in corporate/private organization constituting (15 % of the [Bhutanese samples](#)[Bhutanese sample](#), 5 % of the entire samples), 5 [Indian samples](#)[Indian sample](#) are in the same category which is 10% of the [Indian samples](#)[Indian sample](#), 4 % of the entire samples and 5 are from [Thai samples](#)[Thai sample](#) which is 10% of the [Thai samples](#)[Thai sample](#) and 4% of the entire samples.

Similarly, the respondents working in the Government occupation are 19 respondents from Bhutan (consisting of 49% of [Bhutanese samples](#)[Bhutanese sample](#) and 13% of the entire samples), 5 respondents from India (containing 10% of [Indian samples](#)[Indian sample](#) and 4% of the entire samples) and 2 respondents from Thailand 4% of the [Thai samples](#)[Thai sample](#) and 1% of the entire samples).

The samples in the Business occupation are 3 persons from Bhutan (6% of the [Bhutanese samples](#)[Bhutanese sample](#), 2% of the entire samples), 8 persons from India (17% of [Indian samples](#)[Indian sample](#), 6% from the entire samples), and 3 persons from Thailand (6% of the [Thai samples](#)[Thai sample](#), 2% of the entire samples)

The samples working in the area of Information Technology are constituted of 5 from [Bhutanese samples](#)[Bhutanese sample](#) (10% of the [Bhutanese samples](#)[Bhutanese sample](#),

4% of the entire samples), 4 from Indian (8% of the Indian samples, 3% of entire samples) and 11 from Thai samples (23% of Thai samples, 8% of the entire samples).

Students consist of 11 from Bhutan samples (23% of Bhutanese samples, 8% of entire samples), 26 from India (54% of ~~Indian samples~~ Indian sample, 18% of entire samples) and 25 from Thailand (52% of Thai samples, 17% of entire samples).

Others consist of 3 from Bhutanese samples which 6% of the samples and, 2% of the entire samples, and 2 from Thai (4% of the Thai samples and 1% of the entire samples).

Therefore, in a nutshell the samples' s occupation constitutes of 4% from others, 10% from Business, 12% from Corporation and private, 14% from Information information Technology technology, 18% from Government and 43% from the Student.

### 3.3 Collecting Technique and Attributes

Two Questionnaires were designed to collect respondents' opinion by Google doc application. Then, the Google doc address was shared with the respondents from Bhutan, India and Thailand through email and social media (Facebook). The online survey questionnaires contain 23 questions, first 4 questions were related to the demographic information; the next 19 questions were related to the design of the Text-based CAPTCHA.

The demographic questions gather information on the following information.

- Age
- Gender
- Nationality
- Occupation

The following questionnaires were drawn to collect demographic information and the questions are multiple choice question.

- i. What is your Nationality? This question contains three answers (Bhutanese, Indian and Thai); users are to select any one of the Nationality. This question was used to collect data related to Nationality of the respondents.
- ii. Gender: This question comprises of two options (Male and Female) for users to select. This question was used to capture the user's gender details.
- iii. Age: This question was used for gathering user's age details. The question has 4 choices of age-groups: 10-25, 26-40, 41-60 and 61 and above. The users are required to choose the age-group appropriately.
- iv. Occupation: This question was used to gather information related to user's occupation. The question has 7 choices: Student, Business, Corporation/private organization, Government, Housewife/retiree, Information Technology, and Others. The users are to select the option that best describe their current job.

The other 19 questions ask for respondent's opinion on the usability factors.

These factors are as listed below.

- Awareness and Understanding of CAPTCHA
- Experiences and CAPTCHA perception
- Willingness to use CAPTCHA
- Difficulties of CAPTCHA
- Internet usage frequency
- The CAPTCHA design style and its color.

The questionnaires contain both open-ended and close-ended questions.

Respondents were encouraged to take part in the survey by convincing that participation in this survey would not consume much of their time.

In order to collect data with regards to the usefulness of CAPTCHA, the following questionnaires were drawn.

- Awareness and understanding of CAPTCHA

- Do you know what the Text-based CAPTCHA is? This question wants to measure the knowledge on the characteristics of the Text-based CAPTCHA.
- What do you think CAPTCHA is all about? This question wants to measure the understanding towards the characteristics of CAPTCHA.
- Do you agree that having good CAPTCHA knowledge can help solve CAPTCHA better? : To measure the user's opinion on CAPTCHA Knowledge.

— Experiences and CAPTCHA perception

- Have you ever used the Text-based CAPTCHA? This question wants to detect the familiarity or experiences of users.
- How frequently do you come across CAPTCHA? This question measures the user's experiences with CAPTCHA.
- What types of CAPTCHA are they most familiar with? This question measures the appreciation of users based on their experiences.
- Do you think CAPTCHA is useful? : This question measures the objective of using CAPTCHA.
- What do you think will happen if CAPTCHA is not there? This question wants to gain the user's opinion towards the impact of CAPTCHA.

— Willingness to use CAPTCHA

- Do you accept CAPTCHA to be used? : To measure the user's willingness to used CAPTCHA.
- Do you like using CAPTCHA? This question measures the personal appreciation and difficulties.

- Difficulties of CAPTCHA
  - Is it difficult to identify CAPTCHA? This question measures the opinion based on the user's skill.
  - What makes CAPTCHA identification difficult for you? This question aims to obtain the user's opinion towards attributes of CAPTCHA.
  
- Internet usage frequency
  - Do you agree that you are frequent Internet user? This question wants to measure the experiences in browsing the Internet.
  
- The CAPTCHA design style and its color.
  - Which style of the Text-based CAPTCHA do you prefer?: This question has objective to know the preference of the user towards Text-based CAPTCHA
  - Which color style do you prefer?: This question has objective to find out user's preferences toward the color of the CAPTCHA.
  
- General views of the Respondents
  - Do you think different CAPTCHA should be generated according to age group? : This question has objective to indicate that under the users' point of view, the difficulty in using CAPTCHA may related to age ranges.
  - Do you think CAPTCHA is driving away disable people from internet usage? : This measures the opinion of users for social caring.
  - Please share your views and comments on CAPTCHA? : This collects views of the users with regards to CAPTCHA.

### 3.4 Data analysis

Total of 280 respondents were collected from the online questionnaire which includes the respondents from three different nationalities as mentioned above. These data are analyzed using SPSS v.17 provided by Chulalongkorn University. The objectives of this analysis are to measure the usability of the Text-based CAPTCHA based on users' opinions and suggestions from three nationalities. Moreover, comparisons between these groups as performed in various perspectives are presented in Chapter 4. The statistical analysis in this paper is performed using the categories analysis based on the cross tabular data, and Chi-square with 95% confident interval, significant level is 0.05 ( $\alpha=0.05$ ).

# Chapter IV

## RESULTS

The chapter describes some of the analysis reports which were processed by SPSS v.17. The analysis results are detailed in categories of countries-wise and general wise. Moreover, impacts of the demographic information like Age, Gender, and Occupation were also analyzed. In addition, impacts of experiences and knowledge are discussed. Furthermore, impacts of the Internet browsing are also studied. The details of analysis are narrated in the following sub-topics.

### 4.1 General comparisons among countries

~~The some~~ Some of the main factors of the study are compared and ~~are~~ stated in the following manner. In order to analyzed, the hypothesis were framed and the hypothesis were verified by analyzing the collected survey data.

$H_0$ : The numbers of the ~~Internet-internet~~ users among three countries are the same.

$H_1$ : The numbers of the ~~Internet-internet~~ users among three countries are not the same.

The result indicates that there is a significant difference in the numbers of ~~Internet-internet~~ users among the three countries at  $p\text{-value}=0.000 < \alpha$ . Moreover, when data compared amongst the three countries, it is found that 73% of the Bhutanese samples are ~~Internet-internet~~ users, while 52% of Indian samples are frequent ~~Internet-internet~~ users and when the Thai samples are considered, ~~it is found~~ the result has shown that 100% of the samples are found to be the frequent ~~Internet-internet~~ users as shown in the Table 4.1. Therefore, the numbers of the ~~Internet-internet~~ users among three countries are not the same.

Referring to Table 4.1, it is known that Thai nationalities have the highest ~~Internet-internet~~ accessibility ~~and~~; hence, Thai Nationalities might have better ~~the~~ understandings of CAPTCHA's usefulness.

Table 4.1: Showing Frequency of Internet Users in Percent

	Bhutanese	Indian	Thai	Total
Strongly Disagree	4	0	0	4
Disagree	2	10	0	13
Neither Agree nor Disagree	21	38	0	58
Agree	40	25	33	98
Strongly Agree	33	27	67	127

$H_0$ : The numbers of the ~~Internet-internet~~ users under different ~~Ageage~~-groups are the same.

$H_1$ : The numbers of the ~~Internet-internet~~ users under different ~~Ageage~~-groups are not the same.

The ~~Internet-internet~~ users are compared under different ~~Ageage~~-groups and ~~the result found indicates~~ that there is a significant difference among the different ~~age~~-groups at  $p\text{-value} = 0.00 < \alpha$ . When the age-groups was considered, ~~it was found that the~~there is a subtle difference between the age-groups 10-25 and 26-40 ~~was not much~~; however, it was found a significant difference of users in the age-group 41-60 and the other two groups ~~for m~~More details ~~on figures~~ refer to Table B2 ~~and~~ in the appendix and Table 4.2 . Therefore, the numbers of the ~~Internet-internet~~ users under different ~~Ageage~~-groups are not the same.

Table 4.2 Showing ~~frequency-Frequency~~ of Internet ~~users-Users~~ between ~~ageAge~~-group in ~~percentPercent~~

Age	10-25	26-40	41-60
Strongly Disagree	0	1	9
Disagree	0	6	9
Neither Agree nor Disagree	5	23	64
Agree	35	36	0
Strongly Agree	60	33	18

The Table 4.2 shows the detailed of the relative comparison between age-group with regards to frequency of Internet-internet Usersusers. While-cConsidering the age-groups, while the people in the age range of 26-40 are active users in Bhutanese samples and, respondents in the age-group 10-25 were active users in cases of Thai and Indian. The frequent users consist of 38% (55) of Ageage-group 10-25, 54 % (78) from the age-group 26-40 and 8% (11) from the age-group 41-60.

$H_0$ : The impact based on the gender over the Internet-internet users is the same.

$H_1$ : The impact based on the gender over the Internet-internet users is not the same.

It is also notice that male participants of the samples have more browsing experiences; however, there is not much significant difference. The sample consists of 69 male respondents, out of which 81% agrees to be frequent Internet-internet users. While 69% of the female samples contained 75 women agree to be frequent Internet-internet User-user and the samples contains 75 female. Details of the relative percentage comparison between genders of the sample and its distribution are shown in the Table 4.3. From the table, itThe result can be derive that 75% of the total samples are frequent Internet-internet users.

Table 4.3 Comparison between Genders on frequency-Frequency of Internet usersUsers.

	Male	Female	Total
Strongly Disagree	1	1	2
Disagree	4	2	6

Neither Agree nor Disagree	8	20	28
Agree	28	19	47
Strongly Agree	28	33	61
Total	69	75	144

When the gender was considered for comparison, it is found that there is a significant difference of  $p\text{-value}=0.1030>0.05$ . ~~It is also found that~~Moreover, there is not much difference in internet users. ~~It is found that~~Furthermore, out of 75 females-samples, 52 users are found to be frequent internet user and male at 56 samples out of 69 male samples. Consequently, the impact based on the gender over the ~~Internet-internet~~ users is the same.

$H_1$ : The numbers of ~~Internet-internet~~ users under different occupations are the same.

$H_1$ : The numbers of ~~Internet-internet~~ users under different occupations are not the same.

The result of chi-square shows that there is a significant difference at  $p\text{-value}=0.007<\alpha$ . From the Table B4 ~~in the appendix~~, it found that the samples distribution among the occupations are in the following manner: 12%(17) from corporate, 18%(26) from government, 10%(14) from ~~Businessbusiness~~, 14%(20) from ~~Information-information Technologytechnology~~, 4% (5) from others, and 43%(62) from student. It is found that students constitutes 35% of total samples as frequent ~~Internet-internet~~ users while other categories of occupations are in the range less than 13% of the total samples and some are ~~as-lower as-than~~ 3% of ~~the~~ total samples. Therefore, there is a significant difference of the ~~Internet-internet~~ users under different occupations. In addition, the numbers of ~~Internet-internet~~ users under different occupations are not the same.

## 4.2 The Usefulness Analysis

$H_0$ : The appreciation in using CAPTCHA in three countries is not the same.

$H_1$ : The appreciation in using CAPTCHA in three countries is the same.

The analyzed data shows that the data distributions among the attributes of appreciation have no adequate frequency to conduct the chi-square test. Therefore, the data from some columns are added to obtain the suitable calculation values. As a consequent, the chi-square test can be performed and the result shows that there is a significant difference of appreciations in using CAPTCHA among these three countries at  $p\text{-value}=0.00 < \alpha$ . In addition, the original data were computed and found shown that 54% of Bhutanese samples appreciates, and 76.8 % of the Indian samples do appreciates, while there is only 20% of Thai samples appreciating the usage of CAPTCHA. Therefore, the appreciation in using CAPTCHA in the three countries is not the same.

$H_0$ : The familiarity in using different types of CAPTCHA in three countries is the same.

$H_1$ : The familiarity in using different types of CAPTCHA in three countries is not the same.

The familiarity of using different types of CAPTCHA is seen to be significantly difference different at the  $p\text{-value}=0.00 < \alpha$ . More Details on the familiarity of the using CAPTCHA is shown in the following Table 4.5.

Table 4.3-5 Percentages of Familiarity in Types of CAPTCHA based Nationalities.

Nationality	Text-based CAPTCHA	Image-based CAPTCHA	Audio-based CAPTCHA	Video-based CAPTCHA
Bhutanese	20	6	2	3
Indian	13	12	7	1
Thai	24	9	2	0
Total	57	27	11	4

From the table above, the result indicates that three nationalities are very much familiar with the Text-based CAPTCHA because, 57% of the samples are familiar with the Text-based CAPTCHA, The familiarity of using the other types of CAPTCHA are while 27% with the Image-based CAPTCHA, 11% with the Audio CAPTCHA and 4% with the video-based. Therefore, the familiarity in using different types of CAPTCHA in three countries is not the same.

$H_0$ : The understanding about the usage of CAPTCHA in three countries is the same.

$H_1$ : The understanding about the usage of CAPTCHA in three countries is not the same.

The understanding about the usage of CAPTCHA compared based on Nationalitynationality, Ageage, gender and occupation were considered. When the analysis is performed usingconsider age-ranges, genders and occupations, it is found that the understanding about the usage of the CAPCTHA is not same at p-value < 0.05 (age=0.003, Gender=0.0034, occupation =0.000). However, when the analysis is performed at the National-national level, it is found to haveexpresses that there is the same understanding about the usage of CAPTCHA at p-value=0.1207.

Referring to Table B22A and Table B22, these tables-results indicate that the respondents have different understanding about CAPTCHA. Though, most of the respondents understand that it is useful for information protection and privacy, however, some respondents understands think that CAPTCHA as-is difficult and time consuming.

$H_0$ : The experiences in using CAPTCHA in three countries are the same.

$H_1$ : The experiences in using CAPTCHA in three countries are not the same.

The result shows that the most respondents have experienced in using CAPTCHA as shown in Table 4.58. However, the chi-square analysis cannot be performed because some cells in the table have value less than 5.

Table 4.5-8 Frequency of ~~experieneing~~ Experiencing CAPTCHA

Frequency	Bhutanese	Indian	Thai
Sometimes	32	30	22
Once in a month	4	7	3
once in a week	9	8	17
Daily	3	3	6

When the analysis is performed by taking ~~genders~~, ~~occupations~~ and ~~age-ranges~~, it was found that there are signification differences in the experience at  $p\text{-value} < 0.05$  ( $\alpha$ ) (gender=0.008, occupation=0.008, age=0.054). However, at the national level there is ~~a~~ significant difference in experiences at  $p\text{-vale}=0.140 > 0.05$  ( $\alpha$ ).

$H_0$ : The knowledge in using CAPTCHA within three countries is the same.

$H_1$ : The knowledge in using CAPTCHA within three countries is not the same.

The result shows that there is ~~a~~ significant difference in the knowledge when ~~the~~ comparisons are based on nationality, gender and occupation at  $p\text{-value} < 0.05$  ( $\alpha$ ) (nationality=0.013, Gender=0.032 and occupation=0.025). Thus, it indicates ~~to~~ the knowledge ~~fluctuates-fluctuation when-by~~ comparing with gender, occupation and ~~Nationality~~nationality. However, when the age is considered, the knowledge of the users is found to be ~~the~~ same.

From the Table 4.69, it is clear that Thai samples ~~have-has~~ better knowledge on CAPTCHA with 19% of the users falling in the excellent ~~Category~~category of knowledge.

Table 4.6-9 Frequency ~~table-showing~~ of ~~experieneing~~ Knowledge in CAPTCHA among countries

	Bhutanese	Indian	Thai	Total
Very Little	29	34	19	82
Good	13	6	10	29
Very Good	4	5	10	19
Excellent	2	3	9	14

$H_0$ : The attitude towards the use of CAPTCHA in three countries is the same.

$H_1$ : The attitude towards the use of CAPTCHA in three countries is not the same.

Nevertheless, most people agree that the presented CAPTCHA over the Internet is difficult to type and take time consuming, some people say-said it is boring, while some other say-said it is fun, and somebody say-said it is useful in protecting information as shown in the table 4.7-10. Moreover, the respondents are aware of the use of the CAPTCHA, as because most of the respondents feels that it is helpful in protecting information and privacy.

Table 4.7-10 Users' Attitudes towards CAPTCHA

Nationality	Bhutanese	Indian	Thai
Boring	10	7	3
Time consuming	21	22	10
Difficult to type correctly	28	37	28
Cannot protect information	18	13	14
Can protect information	13	15	38
Fun in typing CAPTCHA	10	7	7
<u>Total</u>	100	100	100

Therefore, the attitude towards CAPTCHA is not same and as we know that there is a significant difference in the liking toward-for the use of CAPTCHA,  $p$ -value=0.000< $\alpha$ . But the majority of the users prefer using-to use it and the majority of the respondents agrees to use further although their choices of CAPTCHA Styles-styles and Colors colors differ significantly. Almost about 75% of the respondents have the willingness to use CAPTCHA.

$H_0$ : The impact of CAPTCHA Styles in usefulness of CAPTCHA is the same.

H<sub>1</sub>: The impact of CAPTCHA Styles in usefulness of CAPTCHA is not the same.

Table 4.8-11 ~~impact~~The Impact of CAPTCHA Style on Usefulness

Impact of CAPTCHA Style on Usefulness of CAPTCHA	
CAPTCHA Style	user's choice
Number(0-9) only without noise and tilt	18
Alphanumeric(A-Z and 0-9) without noise and tilt	19
Number(0-9) only with little noise and tilt	10
Alphanumeric(A-Z and 0-9) with little noise and tilt	15
Alphanumeric(A-Z and 0-9) with high noise and tilt	3
Alphanumeric(A-Z and 0-9) with some overlapping, high noise and large tilt	5
Alphabetic (A-Z) only with little noise and tilt	6
Alphabetic (A-Z) only without noise and tilt	24
Total	100

Referring to Table 4.11, the result is noticed that the choices of users are in variants using careful study of the choices, the CAPTCHA style and user's choices. In addition, it is noticed describes that the majority of the users have made the CAPTCHA Style-style choices which are easy to identify. For example, the highest choices of the users is reflected as "Alphabetic (A-Z) only without noise and tilt" which is theoretically supposed to be easy for identifying as there is no noise and tilt. Thus, from According to the result in the table, the choices differences are in-significant.

Therefore, there is significant impact of CAPTCHA Styles-styles in usefulness of CAPTCHA.

H<sub>0</sub>: CAPTCHA can disrupt disable persons from the Internet-internet usages.

H<sub>1</sub>: CAPTCHA cannot disrupt disable persons from the Internet-internet usages.

Although these samples are normal people, as many as 32% of the samples are not sure that using CAPTCHA may have effects to visually impaired persons. However 46% of the entire population thinks that CAPTCHA can affect the visually impaired persons from using internet, while but 21% of the sample disagree that. Their views are proceeded by using SPSS, the result indicates that views are quite deferent when Nationality and Age was considered at p-value  $< 0.05$  ( $\alpha$ ) (nationality=0.003, Age=0.014) which is less than the significant level.

However, when gender and occupation were considered, it was found that their significant differences are at p-value  $> 0.05$  ( $\alpha$ ) (gender=0.108 and occupation=0.327) which is more than significant level and so there is no effect of indicates it does not affect the visually impaired persons from using internet. All results are shown in Table 4.12.

Table 4.8-12 Frequency indicating Agree and Disagree

	Agree	Neutral	Disagree	Total
Nationality	60	46	30	136
Gender	68	46	30	144
Age	68	46	30	144
Occupation	68	46	30	144
<b>Total</b>	264	184	120	568

$H_0$ : CAPTCHA should be the same for all age groups.

$H_1$ : CAPTCHA should not be the same for all age groups.

The users feel that the CAPTCHA should be generated based on the age group and there is a significant difference with Chi-square test, p-value=0.006 $<$  $\alpha$ , when compared at the national level, gender and occupation are considered, the p-value are 0.122, when gender is considered value=0.444, and when age is considered, and p-

value=0.429 ~~respectively when occupation is considered~~. Over 40% of the respondents are in favor of generating CAPTCHA based on ~~the~~ age-group. When ~~it is compared taking the~~ age, gender and occupation ~~are compared~~, ~~the result in Table 4.13 displays that~~ there is no significant difference. ~~Therefore, it -indicating indicates~~ that CAPTCHA should be generated as per the age group.

Table 4.9-13 Frequency ~~indicating~~ Indicating Aagree and Ddisagree

	Disagree	Neutral	Agree	Total
Nationality	36	50	58	144
Gender	36	50	58	144
Age	36	50	58	144
Occupation	36	50	58	144
Total	144	200	232	576

$H_0$ : CAPTCHA familiarity cannot have any effects in determining the usefulness of CAPTCHA.

$H_1$ : CAPTCHA familiarity can effects in determining the usefulness of CAPTCHA.

Table 4.10-14 Frequency of CAPTCHA Familiarity

Type of CAPTCHA familiar		
	Frequency	Percent
Text-based CAPTCHA	98	68.1
Image-based CAPTCHA	47	32.6
Audio-based CAPTCHA	20	13.9
Video-based CAPTCHA	8	5.6
Total	173	120.1

From the table ~~above~~4.14, it is seen that text-based CAPTCHA has been selected to be the most familiar CAPTCHA. ~~It is seen that because~~ nearly 70% of the total samples are familiar with text-based CAPTCHA. ~~familiarity-Familiarity have-has impacted~~ on the usefulness of CAPTCHA-, ~~as-so~~ it may be easier for the user to use it ~~and-; however~~ the users may be aware of how it is working and what the user have to perform. ~~Therefore, the efficiency of users' usages can be increased. This may increase efficiency of the user in using and therefore familiarity might have impact on the usefulness of CAPTCHA.~~

$H_0$ : CAPTCHA Styles have no impact on CAPTCHA identification.

$H_1$ : CAPTCHA Styles have impact on CAPTCHA identification

The result indicates that there is a significant impact of CAPTCHA Style-style on the ~~Identification-identification~~ of CAPTCHA. Alphabetic (A-Z) only without noise and tilt has been selected by more than 24% of 144 respondents. 3% of the 144 respondents have selected the Alphanumeric (A-Z and 0-9) with high noise and tilt which supposed to be tough CAPTCHA Style. Thus, we can see significant differences in the selection and are depicted in the Table B18 and Table B24. It has been observed that the significant difference among the Nationalities is at  $p\text{-value}=0.006 < \alpha$ .

From the study of two related information contained in the Table B24 and Table B26. It is found that respondents have rated Distorted Character, Confusing Character, Arch, dots and Line interfering as the top contributing factors towards CAPTCHA being problematic in identification of CAPTCHA. The above factors were selected in the followed percentage (29.9%, 34.7%, and 32.7% respectively). And Alphanumeric (A-Z and 0-9) with high noise and tilt from Table B24 indicates that there is impact of noise in the identification of CAPTCHA, only 2.8% of the total sample opted for it, indicating that less people likes high noise and tilt.

$H_0$ : CAPTCHA colors have no impact on CAPTCHA identification.

$H_1$ : CAPTCHA colors have impact on CAPTCHA identification.

The result shows that there is significant difference of  $p\text{-value} = 0.219 > \alpha$  when the age of the users are considered and  $p\text{-value}$  (Nationality=0.334, Gender=0.626 and occupation=0.8810). Thus indicating that there is no significant difference among the Nationalities, Gender, age and Occupation. Multiple color background with single color characters have been selected least indicating that this CAPTCHA color might be a factor determining the usefulness of the CAPTCHA. The details choice of CAPTCHA color are reflected in the Table B25.

$H_0$ : CAPTCHA solving time has no impact on the use of CAPTCHA.

$H_1$ : CAPTCHA solving time has impact on the use of CAPTCHA.

The result indicates that there is significant impact to the user for CAPTCHA being a time consumer. From going through the frequency tables, it is observed that 17.46% of the entire samples feels that CAPTCHA is Difficult and Time Consuming. Moreover, 18.43% of the entire sample thinks that it is time consuming, 12.59% of the entire samples don't like CAPTCHA as user feels it is time consuming. Thus, on average 16.16% of the samples have effect of time consuming. (16.16% of 144=24 persons). Therefore it have some impact on the CAPTCHA uses, however the impact degree could not be properly measured.

$H_0$ : Types of CAPTCHA have no impact on the use of CAPTCHA.

$H_1$ : Types of CAPTCHA have some impacts on the use of CAPTCHA.

When the three nationalities were compared, there was significant difference at  $p\text{-value} = 0.000 < \alpha$  at national level. The Table 4.4 indicates that 57% of the total samples have affinity towards Text-based CAPTCHA and the difference in the choices are seen significant.

Moreover, the table 4.10 displays that different nationalities have significant difference in the choice of CAPTCHA Style The table articulates that 57% of Bhutanese sample opt to use Text-based CAPTCHA, similarly 40% of ~~Indian samples~~ Indian sample and 68% of Thai sample opt to use Text-based CAPTCHA. Therefore, Text-based CAPTCHA is observed to

be more user friendly than the other CAPTCHA. Consequently, it clear that CAPTCHA Style does have impact on the usefulness of CAPTCHA.

$H_0$ : The knowledge about CAPTCHA relates to the CAPTCHA usage.

$H_1$ : The knowledge about CAPTCHA relates to the CAPTCHA usage.

The result also indicates that there is significant difference of users with varying CAPTCHA knowledge at  $p\text{-value} = 0.030 < \alpha$ . The impact is seen at National level but there is no significant difference when the age, gender and occupation of users are considered. It is observed that there are more than 58% of the total sample agrees with. However when nationalities are taken into account, 63% of Bhutanese sample, 33% of Indian sample and 66% of the ~~Thai samples~~Thai sample agrees that it has impact on the usefulness of CAPTCHA.

$H_0$ : The understanding in the use of CAPTCHA between genders is the same.

$H_1$ : The understanding in the use of CAPTCHA between genders is not the same.

And

$H_0$ : The understanding in the use of CAPTCHA between countries is the same.

$H_1$ : The understanding in the use of CAPTCHA between countries is not the same.

And

$H_0$ : The understanding in the use of CAPTCHA between occupations is the same.

$H_1$ : The understanding in the use of CAPTCHA between occupations is not the same.

And

$H_0$ : The understanding in the use of CAPTCHA between age-ranges is the same.

$H_1$ : The understanding in the use of CAPTCHA between age-ranges is not the same.

Furthermore, the analyzed data indicates that males prefer using CAPTCHA more than females. Additionally, males and females have been seen to have same concept about the CAPTCHA usage as indicated in the Table B22. 34% of male sample and 39% of the female sample are aware that CAPTCHA is used for Information protection.

### 4.3 Summarization of the usefulness

The result shows that there are significant impact on the use of CAPTCHA. There is significant difference in the use of Internet amongst the three samples countries. It also depend on the type of CAPTCHA. It is also observed that the CAPTCHA usefulness depends on the knowledge. It is also observed that countries that have better Internet access have better opportunity to know CAPTCHA and have better affinity towards CAPTCHA. In other word the popularity of the CAPTCHA could be achieved through the education of its purposes. Therefore, knowledge plays an important role in determining the usefulness of the CAPTCHA.

Moreover, the development state of the country have impact on determining the usefulness of the CAPTCHA. It is found that addition of noise have impact on the use of the CAPTCHA. Moreover, perceive difficulties plays an important role in determining the usefulness of CAPTCHA. It is found that the CAPTCHA style has impact on the determination of the usefulness of the CAPTCHA.

Furthermore, it is notice that time required to solved the CAPTCHA have impact on the use of CAPTCHA. In addition the CATCHA color does not have significant impact in determining the usefulness of CAPTCHA. The most desired CAPTCHA is Text-based CAPTCHA.

# Chapter V

## DISCUSSION AND CONCLUSIONS

This Chapter is divided into two sections: Discussion and Conclusion. The chapter will discuss issues related to CAPTCHA and its design factors, and their impacts on the usefulness. Moreover, it will discuss the impacts of knowledge, experiences and CAPTCHA exposure and availability of the ~~Internet~~-internet connection as contributing factors towards determining of usefulness of the Text-based CAPTCHA. Based on the discussion and result, the conclusion is drawn.

### 5.1 Discussion

The results show that CAPTCHA designer has to consider many factors in order to come up with design which is more user focus. Some of the factors are CAPTCHA design factors (color, noise, distortion, tilting), knowledge and experiences, age-group, gender and occupations.

The result indicated that major of the respondents believes that the CAPTCHA should be generated as per the age-group. However, some of the respondents have shared their thought and views on this regards. The respondents are concerned about authentication of individual's age if CAPTCHA were to be generated based on the age-group. The respondent thinks that many users may not reveal the correct age in order to have easier CAPTCHA. So there is chances that people may manipulated their age in order to get easier access by solving easy CAPTCHA

According to study [42] of suitability of Text-based CAPTCHA shows that current Text-based CAPTCHA is not suitable for the Age-group of 5-12 years and 60 years and above and the result shows that CAPTCHA should be generated based on the Age-group, which agrees with each other. However, some of the respondents from the sample feels that by generating CAPTCHA based on age group would affect the security effectiveness.

Moreover an Indian respondents believes that CAPTCHA is useful but, the users thinks that there are better alternatives to text based CAPTCHA. A simple logical puzzle (match a shape, drag an object to its description, etc.) is much easier to do (using just mouse, without having to type on a keyboard) and would be quite effective. From this issues, it clear that input method of the CAPTCHA might have influences on determining the usefulness of CAPTCHA.

This Bhutanese sample view, “Sometimes, I feel it boring to typing CAPTCHA. For now, I get used to typing it” also backs the above statement of requiring change in method of input. However, it also indicate that with the gain in experience peoples can adopt the usefulness. Thus indicates that experience have impact on the usefulness of CAPTCHA.

“It’s time consuming yet important too” indicates that time required to solve the CAPTCHA matters and the study also shows that 15% of the sample have same view. Therefore, it essential that the CAPTCHA designer should consider this factors when designing it.

One of the view was, “It's hard to guess between O and 0 in some fonts” as shown in appendix D, it shows that the users have experiences on CAPTCHA and might have encountered such problems. Therefore, it is a part of CATCHA Style which is seen to have impact on the usefulness of CAPTCHA. Consequently it is important to consider the fonts and avoid using characters which are confusing in CAPTCHA.

From the result it is clear that CAPTCHA can affect visually impaired persons from using Internet, however some users have feeling that CAPTCHA should affect the visually impaired persons. The user feels that the existences of “JAWS” software can help visually impaired to used Internet. conversely, a study conducted by Graig [44] reveals that there is problem in using “JAWS” software and found that it is not efficient, as CAPTCHA solving success rate recorded was less than 50% .

This view ,“CAPTCHA sometimes irritates you if more distorted” help in confirming that addition of noise to the CAPTCHA can impact the usefulness of CAPTCHA in align with the study result which indicates that noise have Impact on the usefulness of CAPTCHA. Similarly there is another feedbacks from the users,” It is quite difficult to use and time consuming when we are not able to identify the alphabets and numbers.” Which indicates

that distortion have effect on usefulness of CAPTCHA and the study also shows that 16.8% of the samples thinks that it is difficult to type CAPTCHA correctly. Moreover 32.1% of the sample considers that the distortion is hampering the users in solving CAPTCHA Test.

The view, “It is good....n it protects our privacy n important information” indicates that this user have good knowledge with and the survey result also shows that 58.2% of the total sample have good knowledge on CAPTCHA as depicted on appendix B: Table B2 and Moreover the users regarded CAPTCHA as information defender and almost 25% of the total sample have the same opinion.

The usefulness of the CAPTCHA varies from the country to country due to the development of the Information Technology. With the technology development the knowledge on the CAPTCHA increases and as a consequence, users will developed the knowledge on the usefulness of the CAPTCHA. Thus, the user may use the CAPTCHA.

## **5.2 Conclusions**

Since CAPTCHA is very important technique for computer protection over the network, the use of this technique will be affective only if the presented CAPTCHA is readable or understandable. Therefore, the usability is an issue for CAPTCHA designers before it is implemented. In this study, the results have shown that using CAPTCHA will be useful for some groups of users. Moreover, the effectiveness of using this technique is also relied on the understanding and opportunity of experiencing it. In addition, there is a similarity within these three nationalities in the idea of using different CAPTCHA’s styles in different age groups. Furthermore, according to the probabilities of interacting with CAPTCHA is different in three nationalities, the concerning of using CAPTCHA in disable persons is not the same. The result indicates that people who are familiar with CAPTCHA can understand the difficulty of this technique towards the disable persons. Based on the results obtained from this survey, the CAPTCHA’s designers have to concern various factors, such as nationalities, Age, Gender and occupation of users, experiences of users, understanding and knowledge that users have in their mind, etc. Besides that, the degree of distortion of CAPTCHA contents, input methods, CAPTCHA style are also some of the factors that need to be considered when designing

CAPTCHA. The result indicates that higher experiences and knowledge on the subject can enhanced the users understanding on the usefulness of CAPTCHA. Consequently, the security of the system based on the use of CAPTCHA will be satisfied.

## REFERENCES

- [1] S.Mohammadi and H. Abbasimehr."A High Level Security Mechanism for Internet Polls," in *Signal Processing Systems (ICSPS)*, pp.V3-101-V103-105, 2010.
- [2] L.Shujun, S.Roland, M.Syed Ali Khayam,M. Asad Usman Khan S.Ahmad-Reza and S. Amier Haider Shah. "Breaking e-Banking CAPTCHAs," in *Proceeding ACSAC '10 Proceedings of the 26th Annual Computer Security Applications Conference*, pp.171-180, 2010
- [3] S.Shirali-Shahreza and A.Movaghar. "A New Anti-Spam Protocol Using CAPTCHA," in *Proceedings of the 2007 IEEE International Conference on Networking, Sensing and Control*, pp. 234 – 238, 2007.
- [4] Liming Wang, Xiuling Chang, Zhongjie Ren , Haichang Gao , Xiyang Liu and U .Aickelin. "Against Spyware Using CAPTCHA in Graphical Password Scheme," in *Advanced Information Networking and Applications (AINA)*, pp. 760 – 767, 2010
- [5] R.P.Karrer."EC: an edge-based architecture against DDoS attacks and malware spread," in *Advanced Information Networking and Applications (AINA)*, pp. 49 – 56, 2006.
- [6] S.Shirali-Shahreza and M. Shirali-Shahreza. "Identifying Child Users: Is It Possible?," in *SICE Annual Conference*, pp. 3241 – 3244,2008.
- [7] S.Shirali-Shahreza and M. Shirali-Shahreza." Multilingual Highlighting CAPTCHA," in *Information Technology: New Generations (ITNG)*, pp. 447 – 452, 2011.
- [8] Persian Arabic Baffletext CAPTCHA.
- [9] Haichang Gao, Honggang Liu, Dan Yao,Xiyang Liu,U.Aickelin. "An audio CAPTCHA to distinguish humans from computers, "in *Electronic Commerce and Security (ISECS)*,pp. 265 – 269,2010.

- [10] Haichang Gao, Dan Yao , Honggang Liu, Xiyang Liu, Liming Wang. "A Novel Image Based CAPTCHA Using Jigsaw Puzzle," in *Computational Science and Engineering (CSE)*, pp. 351 – 356,2010.
- [11] Kuo-Feng Hwang , Cian-Cih Huang and Geeng-Neng You. "A Spelling Based CAPTCHA System by Using Click," in *International Symposium on Biometrics and Security Technologies*, pp. 1-8, 2012.
- [12] M.Shirali-Shahreza and S.Shirali-Shahreza."Question-Based CAPTCHA," in *Conference on Computational Intelligence and Multimedia Applications*, pp. 54-58, 2007.
- [13] M. Blum LVA, and J. Langford, The CAPTCHA Project (Completely Automatic Public Turing Test to tell Computers and Humans Apart). School of Computer Science, Carnegie-Mellon University 2000.
- [14] A.A. Chandavale, A.M.Sapkal and R.M.Jalnekar. "Algorithm To Break Visual pp.262, 2009.
- [15] J.Zhang and X.Wang. "Breaking Internet Banking CAPTCHA Based on Instance Learning," in *International Symposium on Computational Intelligence and Design*, pp. 39-43, 2010.
- [16] T.Yamamoto, T.Suzuki and M.Nishigaki. "A Proposal of Four-Panel Cartoon CAPTCHA," in *International Conference on Advanced Information Networking and Applications*, pp. 159-166, 2011.
- [17] N.Nanglae and P.Bhattarakosol. "A Study of Human Bio-Detection Function under Text-Based CAPTCHA System," in *Computer and Information Science (ICIS)*, pp. 139 – 144, 2012.
- [18] H.Chen-Chi,W.Chen-Chi,C.Kuan-Ta and L.Chin-Laung," DevilTyper: A Game for CAPTCHA Usability Evaluation," *Computers in Entertainment*. vol. 9. New York, pp. 1-14, 2011.
- [19] Bin B. Zhu JY, Qiujie Li, Chao Yang, Jia Liu, Ning Xu, Meng Yi, Kaiwei Cai, "Attacks and Design of Image Recognition CAPTCHAs," *Proceeding CCS '10*

- Proceedings of the 17th ACM conference on Computer and communications security, New York, pp. 187-200, 2010.
- [20] G.Mori and J.Malik,"Recognizing objects in Adversal Clutter: Breaking a visual CAPTCHA," in *Computer Vision and Pattern Recognition*, 2003.
- [21] L.V.Ahn, M.Blum and J.Langford, "Telling Humans and Computers Apart Automatically," in *communications of the ACM*, pp. 57-60, 2004.
- [22] J.Nielsen, "Usability Engineering", pp.24-37, 1993
- [23] K.C. Madathil,G.F. Alapatt and J.S. Greenstein, "An Investigation of the Usability of Image-based CAPTCHAs," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 54 no. 16, pp.1249-1253, 2010.
- [24] H.D. Truong, C.F. Turner and C.C. Zou,"iCAPTCHA: The Next Generation of CAPTCHA Designed to Defend against 3rd Party Human Attacks," *Communications (ICC), 2011 IEEE International Conference*, pp. 1-6, 5-9 June 2011.
- [25] J. Yan and A.S. El Ahmad,"A Low-Cost Attack on a Microsoft CAPTCHA," *proceedings of the 15th ACM conference on computer and communications security*, 2008.
- [26] Sam Hocevar, "PWNtcha - a Captcha Decoder Website", <http://caca.zoy.org/wiki/PWNtcha>
- [27] Dancho Danchev, "Inside India's CAPTCHA Solving Economy", <http://blogs.zdnet.com/security/?p=1835>, 2008
- [28] Yan, Jeff & El Ahmad, Ahmad Salah "Usability of CAPTCHAs or usability issues in CAPTCHA design" in *Proceedings of Symposium on usable privacy and security (SOUPS)*, 2008, pp.10.
- [29] S. Vikram,Y. Fan and G.Gu,"SEMAGE:A new Image-based Two -factor CAPTCHA," *ACM 2011*,pp.237-246, 2011.
- [30] W.AI-Sudani, J. Wang and F. Liu," Protection through Multimedia CAPTCHAS", *proceeding of Mobile Multimedia Security*, pp.63-68, 2010.
- [31] <http://www.captcha.net/> [Accessed 15-03-2013]

- [32] M.Noar,"Verification of s human in the loop, or identification via the Turing test", <http://www.wisdom.weizmann.ac.il/~naor/PAPERS/human.pdf> [Accessed on 15-03-2013]
- [33] [http://en.wikipedia.org/wiki/Communications\\_in\\_India](http://en.wikipedia.org/wiki/Communications_in_India) [Accessed 15-03-2013]
- [34] [http://en.wikipedia.org/wiki/Internet\\_in\\_Thailand](http://en.wikipedia.org/wiki/Internet_in_Thailand) [Accessed 15-03-2013]
- [35] <http://www.internetworldstats.com/stats3.htm> [Accessed 15-03-2013]
- [36] C.A.Fidas,N.M.Avouris and A.G.Voyiatzis, "On the Necessity of User-Friendly CAPTCHA", *Proceedings of ACM Conference on Human Factors in Computing Systems*. Vancouver, BC, Canada, pp. 2623-2626 ACM press, May 7-12, 2011
- [37] Drs. Erik P.W.M,"Questionnaire based usability testing", *conference proceedings European Software Quality Week*, Brussels, pp.9, November 1998.
- [38] K.Chellapilla and P.Simard, "Using machine Learning to Break Visual Human Interaction Proofs (HIPs)," *Advances in Neural Information Processing System (NIPS'2004)*, Vancouver, Canada, MIT Press, December 2004.
- [39] W.Yu, X.Wang, X.Fu, D.Xuan and W.Zhao,"An invisible Localization Attack to Internet Threat Monitors," *IEEE Parallel and Distributed Systems*, vol. 20, no.11, pp. 1611-1625, Nov 2009.
- [40] [http://en.wikipedia.org/wiki/Literacy\\_rates](http://en.wikipedia.org/wiki/Literacy_rates) [Accessed on 27-03-2013]
- [41] <http://www.bbc.co.uk/news/technology-21260007> [Accessed on 27-03-2013]
- [42] R.Pakdel, N.Ithnin and M.Hashemi, "CAPTCHA: A survey of Usability Features," *Research journal of information technology*, pp.215-228, 2011.
- [43] <http://www.indexmundi.com/facts/bhutan/internet-users> (Accessed on 27-03-2013).
- [44] G.Sauer, H.Hochheiser, J.Feng and J.Lazar,"Towards a Universally Usable CAPTCHA", in *Proceedings of the 4th Symposium On Usable Privacy and Security (SOUPS'08)*, pp.4, 2008.

\

## APPENDICES

## APPENDIX A

## IMAGES OF CAPTCHA

Figure A1: A kind of CAPTCHA: Circle aptcha

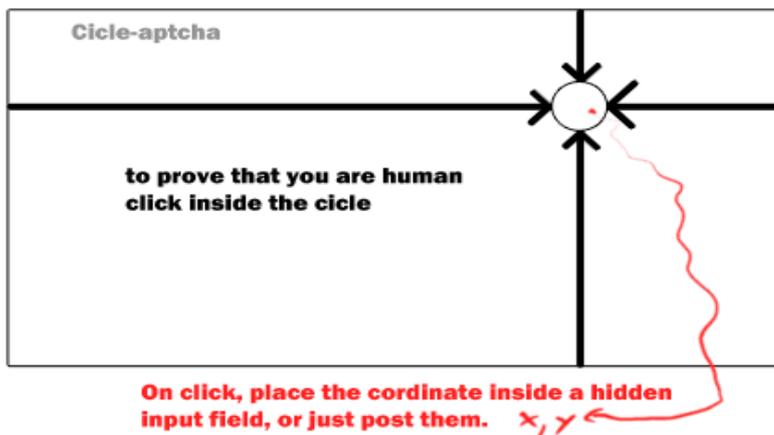


Figure A2: Example Question based CAPTCHA

CAPTCHA

This question is for testing whether you are a human visitor and to prevent automated spam submissions.

**Math question: \***

3 + 19 =

Solve this math question and enter the solution with digits. E.g. for "two plus four = ?" enter "6".

Figure A3: Another Example of Question based CAPTCHA

**Qualifying question**

Just to prove you are a human, please answer the following math challenge.

**Q:** Calculate:

$$\frac{\partial}{\partial x} \left[ 7 \cdot \sin \left( 5 \cdot x - \frac{\pi}{2} \right) + \cos \left( 3 \cdot x + \frac{\pi}{2} \right) \right] \Bigg|_{x=2\pi}$$

**A:**

*mandatory*

Note: If you do not know the answer to this question, reload the page and you'll (probably) get another, easier, question.

## APPENDIX B

## EXPERIMENTAL SAMPLES

Table B1: Sample collected Details

Sample Collected grouped in nationality, Gender and Age					
Age		Nationality	Gender		Total
			Male	Female	
10-25		Bhutanese	17	22	39
		Indian	8	7	15
		Thai	23	22	45
		Total	48	51	99
26-40		Bhutanese	62	44	106
		Indian	6	20	26
		Thai	12	13	25
		Total	80	77	157
41-60		Bhutanese	10	7	17
		Indian	6	1	7
		Total	16	8	24

Table B2: Indicates the CAPTCHA knowledge of the Samples

Knowledge of CAPTCHA				
Nationality	Very Little	Good	Very Good	Excellent
Bhutanese	57%	23%	14%	6%
Indian	71%	13%	6%	6%
Thai	39%	21%	21%	19%

Table B3: The respondents view with regards to CAPTCHA

Nationality	Protect Information & Privacy	Personal Usage	Easy to understand & Type	Difficult & Time consuming	Boring & Useless	Hack by malware
Bhutanese	40%	18%	18%	11%	8%	5%
Indian	39%	17%	33%	9%	2%	0%
Thai	28%	18%	17%	26%	8%	4%
Total	107%	52%	68%	46%	18%	9%

Table B4: How the user gets to know CAPTCHA or Source of CAPTCHA

Nationality					
	Registering E-mail account	while registering form	Using Internet banking	Do not come across CAPTCHA	Do not surf IE
Bhutanese	45%	38%	15%	0%	2%
Indian	36%	45%	18%	0%	0%
Thai	40%	51%	7%	0%	2%
	121%	134%	41%	0%	4%

Table B5: User's selected choice group by nationalities

Type of CAPCHA which user group by Nationalities				
Nationalities	Text-based	Image-based	Audio-based	Video-based
Bhutanese	57%	23%	14%	7%
Indian	40%	36%	21%	3%
Thai	68%	24%	7%	1%

User's Choice of CAPTCHA based on Age of the respondent				
Age	Text-based	Image-based	Audio-based	Video-Based
10-25	76%	31%	15%	4%
26-40	30%	28%	17%	6%
41-60	76%	57%	14%	10%

Table B6: User's selected choice group by Age

CAPTCHA Choice by Gender				
Gender	Text-based	Image-based	Audio-based	Video-based
Male	40%	16%	9%	2%
Female	29%	15%	8%	4%
Total	69%	31%	16%	6%

Table B7: Users' CAPTCHA selection by Gender

Nationality	Frequency of CAPTCHA			
	Sometimes	Monthly	Weekly	Daily
Bhutanese	69%	10%	12%	9%
Indian	63%	15%	17%	6%
Thai	46%	6%	30%	19%

Table B8: frequency of CAPTCHA the respondents confront

Likings and disliking of CAPTCHA									
Nationality	Like using CAPTCHA	like CAPTCHA is enjoyable	like CAPTCHA is like game	like CAPTCHA is Interesting	Don't like CAPTCHA is time consuming	Don't like CAPTCHA is useless	Don't like CAPTCHA content not clear	Don't like CAPTCHA is boring	Don't like CAPTCHA is annoying
Bhutanese	24	6	9	14	17	7	13	5	5
Indian	12	8	29	29	6	6	6	0	6
Thai	3	3	6	11	17	7	24	13	16
	38	17	43	54	39	20	43	19	27

Table B9: Reasons for liking and disliking of CAPTCHA

Nationality	Distorted Character	Similar background & char color	Arch, dots and Line interfering	Confusing Character	Tilting Character	Overlapping Character
Bhutanese	18	14	13	13	9	15
Indian	13	11	29	20	6	9
Thai	17	13	14	21	6	13
Total	47	38	56	55	20	37

Table B10: Influential factors affecting identification of CAPTCHA

The views on usefulness of CAPTCHA						
Nationality	Boring	Time consuming	difficult to type correctly	cannot protect information	can protect information	Fun in typing CAPTCHA
Bhutanese	9%	22%	16%	17%	25%	12%
Indian	7%	22%	37%	13%	15%	8%
Thai	5%	10%	23%	18%	38%	8%
Total	21%	53%	76%	47%	77%	27%

Table B11: Views on the usefulness of CAPTCHA

View of respondent if the CAPTCHA is not there

Nationality	Data hacked easily	Nothing will happen	Computer at Risk	Flexible to use computer	Trust my working system	Feel unsecured
Bhutanese	28%	18%	8%	14%	12%	20%
Indian	19%	22%	22%	15%	11%	11%
Thai	23%	25%	14%	10%	16%	11%
Total	69%	65%	45%	39%	40%	42%

Table B12: The views of users if the CAPTCHA is not there

The CAPTCHA Style preferences

Nationality	Number(0-9) only without noise and tilt	Alphanumeric(A-Z and 0-9) without noise and tilt	Number(0-9) only with little noise and tilt	Alphanumeric(A-Z and 0-9) with little noise and tilt	Alphanumeric(A-Z and 0-9) with high noise and tilt	Alphanumeric(A-Z and 0-9) with some overlapping, high noise and large tilt	Alphabetic (A-Z) only with little noise and tilt	Alphabetic (A-Z) only without noise and tilt
Bhutanese	12	17	9	10	2	1	10	38
Indian	27	17	8	6	0	4	8	29
Thai	14	19	9	26	3	13	3	14
Total	54	52	26	42	5	18	22	81

Table B13: The CAPTCHA Style for choice of users

Nationality	Strongly un-accepted	Partly un-accepted	Neutral	Partly accepted	Strongly Accepted
Bhutanese	7%	11%	32%	25%	25%
Indian	2%	29%	35%	21%	13%
Thai	3%	1%	40%	44%	11%
Total	12%	42%	108%	90%	49%

Table B14: Users' willingness to adopt the CAPTCHA

Nationality	Strongly Disagree	Partly Disagree	Neutral	Partly Agree	Strongly Agree
Bhutanese	5%	7%	25%	38%	25%
Indian	4%	25%	38%	19%	15%
Thai	3%	7%	24%	44%	21%
Total	12%	40%	87%	101%	61%

Table B15: Effect of CAPTCHA knowledge

internet frequent user	Frequency	Percent
Strongly Disagree	3	1.1
Disagree	15	5.4
Neither Agree nor Disagree	43	15.4
Agree	98	35.0
Strongly Agree	121	43.2
Total	280	100.0

Table B16: The frequency details of Internet common user

Difficulties level	CAPTCHA Style								Total
	Number(0-9) only without noise and tilt	Alphanumeric(A-Z and 0-9) without noise and tilt	Number(0-9) only with little noise and tilt	Alphanumeric(A-Z and 0-9) with little noise and tilt	Alphanumeric(A-Z and 0-9) with high noise and tilt	Alphanumeric(A-Z and 0-9) with some overlapping, high noise	Alphabetic (A-Z) only with little noise and tilt	Alphabetic (A-Z) only without noise and tilt	
Very Difficult	9	0	3	1	1	0	3	9	26
Difficult	10	14	7	7	1	1	5	28	73
Neutral	19	29	13	29	3	12	12	37	154
Easy	4	4	1	1	1	0	3	10	24
Very Easy	1	1	0	0	0	0	0	1	3
	43	48	24	38	6	13	23	85	280

Table B18 Effect of CAPTCHA Style on identifying CAPTCHA

Age * internet frequent user Cross tabulation						
Age	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Total
10-25	0	0	2	14	20	99/36%
26-40	1	4	10	21	21	157/57%
41-60	0	1	4	1	2	2/8%
Total	1	5	15	35	43	280
Gender * internet frequent user Cross tabulation						
Gender						
Male	0	4	5	19	23	144/51%
Female	1	2	10	16	20	136/49%
Total	1	5	15	35	43	280
Nationality * internet frequent user Cross tabulation						
Nationality						
Bhutanese	1	4	9	23	21	162/58%
Indian	0	2	6	4	5	48/17%
Thai	0	0	0	8	18	70/26%
Total	1	5	15	35	43	280
Occupation * internet frequent user Cross tabulation						
Occupation						
corporation/private	0	1	3	10	5	53/19%
Government	0	2	4	8	9	62/23%
Business	1	1	4	0	2	21/8%
Information Technology	0	0	0	4	8	31/12%
Others	0	0	0	3	1	10/4%

Student	0	2	4	11	20	103/37%
Total	1	5	15	35	43	280

Table B21: Internet frequency group by Age, Gender, nationality and occupation

Gender	useful for information protection and privacy	useful for personal usage	easy to understand and type	Difficult & Time consuming	Boring & Useless	Hack by malware
Male	58	32	31	30	12	7
Female	57	24	33	17	10	5
	115	56	64	47	22	12

Table B22: Gender impact on CAPTCHA understanding

Know CAPTCHA	Think :Boring	Think :Time consuming	Think : difficult to type correctly	Think :cannot protect information	Think :can protect information	Think :Fun in typing CAPTCHA
Very Little	12	34	25	23	24	14
Good	3	2	13	6	11	3
Very Good	2	4	7	5	12	1
Excellent	0	2	2	1	7	3

Total	17	42	47	35	54	21
-------	----	----	----	----	----	----

Table B23: CAPTCHA knowledge Perceptions

CAPTCHA Style	Frequency	Percent
Number(0-9) only without noise and tilt	43	15.4
Alphanumeric(A-Z and 0-9) without noise and tilt	48	17.1
Number(0-9) only with little noise and tilt	24	8.6
Alphanumeric(A-Z and 0-9) with little noise and tilt	38	13.6
Alphanumeric(A-Z and 0-9) with high noise and tilt	6	2.1
Alphanumeric(A-Z and 0-9) with some overlapping, high noise and large tilt	13	4.6
Alphabetic (A-Z) only with little noise and tilt	23	8.2
Alphabetic (A-Z) only without noise and tilt	85	30.4

Table B24: CAPTCHA Style impact

CAPTCHA Color Style	Frequency	Percent
Single color background with single color characters	132	47.1
Single color background with multiple color characters	63	22.5
Multiple color background with multiple color characters	57	20.4

Multiple color background with single color characters	28	10.0
--	----	------

Table B25: CAPTCHA Color impact

Difficulties in identifying CAPTCHA	Frequency	Percent
Distorted Character	90	32.1
Similar background & character color	72	25.7
Arch, dots and Line interfering	83	29.6
Confusing Character	90	32.1
Tilting Character	39	13.9
Overlapping Character	72	25.7

Table B26 Difficulties in Identifying CAPTCHA

APPENDIX C  
SURVEY QUESTIONNAIRE

1. what is your Nationality ?

- Thai
- Bhutanese
- Indian

2. Gender

- Male
- Female

3. Age: select the age group in which your Age falls in.

- 10-25
- 26-40
- 41-60
- 60 and Above

4. Occupation: select your Occupation.

- Student
- Business
- Corporation/private organization
- Government
- Housewife/retiree
- Information Technology
- Others

5. Do you agree that you are frequent Internet user?

- Agree Strongly
- Agree
- Neither Agree nor Disagree
- Disagree
- Disagree Strongly

6. Do you know what Text-based CAPTCHA is?

- Not at All
- Very Little
- Good
- Very Good
- Excellent

7. How do you feel about CAPTCHA or what do you think CAPTCHA is all about?

- Very useful for information protection and privacy
- Very useful for personal usage
- Very easy to understand and type
- Very difficult to type and time consuming
- Very boring and useless for information protection
- Very easy to hack by malware or malicious software

8. Have you ever used the Text-based CAPTCHA?

- Yes, when registering my new e-mail account
- Yes, when browsing some registration forms
- Yes, when using Internet Banking services
- No, I did not come across with such CAPTCHA
- No, I do not surf Internet much

9. Type of CAPTCHA you are familiar with

- Text-based CAPTCHA
- Image-based CAPTCHA
- Audio-based CAPTCHA
- Video-based CAPTCHA

10. How frequently do you come across CAPTCHA?

- Daily
- Weekly
- Monthly
- Sometimes
- Never come across

11. Do you like Using Text-based CAPTCHA?(You can choose more than one)

- I like using it
- It is enjoyable to use it
- It is like a game
- It is interesting
- It is time consuming
- It is useless
- Its text contents are not clear
- It is boring to use it
- It is annoying to use it

12. Is it difficult to identify the CAPTCHA?

- Very difficult
- Difficult
- Neither Difficult nor Easy
- Easy
- Very Easy

13. What makes CAPTCHA identification difficult for you? (You can choose more than one)

- Having Distorted Characters in the CAPTCHA
- Having Similar background and character Color
- Having arch, dots and lines interfering the CAPTCHA contents
- Having Confusing Characters in the CAPTCHA contents
- Tilting of the Characters in the CAPTCHA
- Joining or overlapping of the Characters in the CAPTCHA

14. Do you think CAPTCHA is useful? (You can choose more than 1)

- It is boring
- It is time consuming
- It is difficult to type correctly
- I believe that it cannot really protect my information
- I believe that it can protect my information well
- It is fun in typing to solve the CAPTCHA

15. If you don't need to enter CAPTCHA, what do you feel? Or what do you think will happen if CAPTCHA is not there? (You can choose more than 1)

- My data will be easily hacked
- Nothing different from having CAPTCHA
- My computer is at risk
- Very flexible to use my computer
- I can still trust my working system
- I feel unsecured

16. Which style of Text-based CAPTCHA do you prefer?

- Alphabetic (A-Z) only without noise and tilt
- Number (0-9) only without noise and tilt
- Alphanumeric (A-Z and 0-9) without noise and tilt

- Alphabetic (A-Z) only with little noise and tilt
- Number (0-9) only with little noise and tilt
- Alphanumeric (A-Z and 0-9) with little noise and tilt
- Alphanumeric (A-Z and 0-9) with high noise and tilt
- Alphanumeric (A-Z and 0-9) with some overlapping, high noise and large tilt

17. Which color style do you prefer?

- Single color background with single color characters
- Single color background with multiple color characters
- Multiple color background with multiple color characters
- Multiple color background with single color characters

18. Do you accept CAPTCHA to be used?

- Accept Strongly
- Accepted
- Neutral
- Un-accepted
- Un-accepted Strongly

19. Do you agree that CAPTCHA should be same for every people?

- Agree Strongly
- Agree
- Neither Agree nor Disagree
- Disagree
- Disagree Strongly

20. Do you agree that age group should be considered in generating CAPTCHA?

- Agree Strongly
- Agree
- Neither Agree nor Disagree
- Disagree

Disagree Strongly

21. Do you agree that Text-based CAPTCHA can bar the visually impaired people from the Internet Usage?

- Agree Strongly
- Agree
- Neither Agree nor Disagree
- Disagree
- Disagree Strongly

22. Do you agree that having good CAPTCHA knowledge can help solve CAPTCHA better?

- Agree Strongly
- Agree
- Neither Agree nor Disagree
- Disagree
- Disagree Strongly

23. Please share your views and comments on CAPTCHA



Submit

## APPENDIX D

### RESPONDENTS' VIEW

With the collection of the Data, some of the respondents view are also collected and found some views interestingly to be noted. The views are categorized into Nationalities. Therefore the views are listed below with classification.

#### I. Indian Views

1. My view about CAPTCHA is a conceptual thinking of securing the data.... Me as a IT professional, CAPTCHA is good way to securing once data since it's very hard to predict how and what will be next CAPTCHA if you do mistake in typing.
2. It's very interesting and liked it
3. Seems it very good to use it...but only thing is i never come across the word CAPTCHA.
4. CAPTCHA can be useful. But I think there are better alternatives to text based CAPTCHA. A simple logical puzzle (match a shape, drag an object to its description, etc.) is much easier to do (using just mouse, without having to type on a keyboard) and would be quite effective.
5. It is fun in typing to solve the CAPTCHA.

#### II. Thai Views

1. Sometimes, I feel it boring to typing CAPTCHA. For now, I get used to typing it.
2. CAPTCHA is not bothering me : D
3. It's hard to guess between O and 0 in some fonts.
4. I think this program will be successful when you make the suggestion of this program easier for everybody even though kids or person who did not know more the information about CAPTCHA/computer by more picture or more multimedia.
5. Interesting
6. Its ok for me, I would like to use it.

### III. Bhutanese Views

1. it's very boring to those who doesn't know and never come across the CAPTCHA and hope it is very interesting and important to those who come across the word CAPTCHA
2. Excellent technology, quite beneficial.
3. It's time consuming yet important too.
4. CAPTCHA, what I think is important that gives identification of a respective person even though s/he has a same name with other person.
5. CAPTCHA is important for securing our internet account and information.
6. Text-based CAPTCHAs are less secure. It should always be combined with text and numeric figures for the less secure data and for high secure data like banking transaction, etc. it should be the combination of image, audio text, numeric and special characters in order to secure that data.
7. I have never experienced other types of CAPTCHA except text-based CAPTCHA. So, I think the most commonly used is this and people prefer this only. Therefore, I don't have much ideas about other types of CAPTCHA.
8. CAPTCHA is time consuming...
9. CAPTCHA sometimes irritates you if more distorted
10. I could say CAPTCHA is very important for the protection.
11. Generating CAPTCHA according to age group can be a bit tricky as it's hard to determine the age of the person who is using CAPTCHA to log in to a particular site.
12. The CAPTCHA is good for securing the information especially for online data and its verification required.
13. CAPTCHA is very important for personal use
14. CAPTCHA is a good idea to keep the information secure.
15. Its program guarantee cannot be trusted

16. It is quite difficult to use and time consuming when we are not able to identify the alphabets and numbers.
17. very boring and don't like using CAPTCHA

## **BIOGRAPHY**

Kuenzang Norbu, an employee at National Pension and Provident fund has graduated a bachelor degree in Computer Application from VMRF Deemed University, Salem, Tamil Nadu. He has been working at National Pension and Provident of Bhutan since 2002 in Information Technology department. Currently, doing master degree in Information Technology and Computer Science at Chulalongkorn University, Thailand, in 2013.