การระบุสิ่งที่มีอิทธิพลอย่างมีนัยสำคัญต่อการทวนสอบ CAPTCHA เชิงข้อความจากลักษณะเฉพาะส่วนบุคคล

นายเทเตน ทามัง

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2555

IDENTIFICATION OF SIGNIFICANT INFLUENCERS ON TEXT-BASED CAPTCHA VERIFICATION FROM PERSONAL CHARACTERISTICS

Mr. Tsheten Tamang

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

Thesis Title	IDENTIFICATION OF SIGNIFICANT INFLUENCERS ON TEXT-BASED CAPTCHA VERIFICATION FROM PERSONAL CHARACTERISTICS
By	Mr. Tsheten Tamang
Field of Study	Computer Science and Information Technology
Thesis Advisor	Assistant Professor Pattarasinee Bhattarakosol, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Master's Degree

.....Dean of the Faculty of Science
(Professor Supot Hannongbua, Dr. rer. nat.)

THESIS COMMITTEE

.....Chairman
(Assistant Professor Nagul Cooharojananone, Ph.D.)

.....Thesis Advisor
(Assistant Professor Pattarasinee Bhattarakosol, Ph.D.)

.....External Examiner
(Assistant Professor Ohm Sornil, Ph.D.)

นายเทเตนทามัง:การระบุสิ่งที่มีอิทธิพลอย่างมีนัยสำคัญต่อการทวนสอบCAPTCHAเชิงข้อความจากลักษณะเฉพาะส่วนบุคคล. (IDENTIFICATION OF SIGNIFICANT INFLUENCERS ON TEXT-BASED CAPTCHA VERIFICATION FROM PERSONAL CHARACTERISTICS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ผศ.ดร.ภัทรสินี ภัทรโกศล, 61 หน้า.

ในปัจจุบันCAPTCHAเป็นเครื่องมือสำคัญที่ช่วยเพิ่มความปลอดภัยในการเข้าสู่ระบบที่ต้องการ อย่างไรก็ตามแม้ว่าผู้ใช้บางคนจะเป็นผู้ใช้ที่ได้รับอนุญาตให้เข้าสู่ระบบก็ยังคงมีความยุ่งยากในการพิมพ์CAPTCHAเกิดขึ้นและเนื่องด้วยCAPTCHAชนิดตัวอักษรเป็นประเภทที่นิยมมากที่สุดในบรรดาชนิดของCAPTCHAทั้งหมดดังนั้นจึงได้ศึกษาในเรื่องความยากของCAPTCHAชนิดตัวอักษรผลการศึกษาแสดงให้เห็นว่าลักษณะของตัวอักษรเพศของผู้ใช้และการศึกษาพื้นฐานของผู้ใช้เป็นตัวแปรสำคัญที่ใช้ชี้วัดความถูกต้องในการพิมพ์CAPTCHAของผู้ใช้ดังนั้นCAPTCHAชนิดตัวอักษรจะต้องมีความเหมาะสมของลักษณะตัวอักษรประกอบกับการศึกษาพื้นฐานและเพศของผู้ใช้เอง

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2555

5472606023 : : MAJOR COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
KEYWORDS : CAPTCHA/TEXT-BASED CAPTCHA/CAPTCHA TEXT/CAPTCHA COLOR

TSHETEN TAMANG: IDENTIFICATION OF SIGNIFICANT INFLUENCERS ON TEXT-BASED CAPTCHA VERIFICATION FROM PERSONAL CHARACTERISTICS. THESIS ADVISOR: ASST. PROF. PATTARASINEE BHATTARAKOSOL, Ph.D., 61 pp.

Presently, CAPTCHA is an important mechanism to gain access to the required system. However, there are some difficulties for users in typing CAPTCHA although they are authorized persons. Since the Text-based CAPTCHA is the most popular mechanism amongst all the CAPTCHA techniques, the difficulties of this Text-based are studied and drawn out. The results of this study have shown that the presented character(s), genders of users, and their educational background are some of the important factors determining the correctness of CAPTCHA typing by its users. Therefore, generating a Text-based CAPTCHA must use the appropriate character(s), that also combining with the educational background and gender of the users.

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

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

Academic Year : 2012

ACKNOWLEDGEMENTS

I am indeed grateful to the Thailand International Development Cooperation Agency (TICA) of Thailand and the National Pension and Provident Fund (NPPF) of Bhutan who supported this research.

This research would not have been possible without the unwavering support of my advisor, Assistant Professor Pattarasinee Bhattarakosol, Ph.D. Therefore, I remain particularly obliged to her for her invaluable contributions through constructive suggestions and recommendations. In addition, I would like to thank the thesis committees, Assistant Professor Nagul Cooharojananone, PhD., and Assistant Professor Ohm Sornil, PhD. for taking their precious time in thesis examinations and for their invaluable comments.

Finally, my graduation would not have been achieved without best wish from my parents and my relation, who helped me in everything and always giving me greatest love, willpower and financial support until completion of my study and, lastly my heartfelt special thanks to my friends and my colleagues in INSET Lab. for their help and encouragement.

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Background and Importance

Presently, living has changed its dimension with the introduction of the internet to mankind, ways people connect to each other, advertising, shopping, education, etc. Consequently, system security has become the most important issue for any websites since there are many methods used to intrude the system over the Internet. Different measures have been developed and adapted to fight this kind of illegal activities which have always been the greatest fear.

There are different types of protection like Physical and Logical protection. The Physical protection for a prohibited access are related to various methods, such as hiring guards, implementation of a Closed Circuit Television (CCTV), implementation of an accessing room policy, etc. However, this protection can protect only when the system is being approached by human. Unfortunately, these physical intrusions are not a serious problem for all computer users unlike the logical invasion. Besides the physical protection, the logical roles are even more vital. The logical protection concerns with the uses of software protection methods that can protect the invasion from malicious software even from a distance location that is trying to access the system.

With increase in usage of computers and everything being saved in the storage devices, this information, so called data, has become very important. Thus, it has to be securely protected from the unwanted access. Various methods are implied these days to secure the data, such as the implementation of firewalls, antivirus, internet securities, etc. However, there is nothing 100% safe, as-the fact that-The unwanted program that tries to intrude the systems are also developed equally as efficient as the security systems; with the rapid competition among the security systems with the unwanted programs, it has always be concerned that all data should be protected.

Many Researchers have proposed many data protection solutions in both physical and logical. This research basically concern towards one of the logical protection so called CAPTCHA ("Completely Automated Public Turing test to tell Computers and Humans Apart"). CAPTCHA is a program that protects websites against bots by generating and grading

test that only humans can pass [1]. The term "CAPTCHA" was originally framed in 2000 by Luis Von Ahn Manuel Blum, Nicholas J. Hopper, and John Langford (all of Carnegie Mellon University) [1]. It was designed to ensure that it is not an attempt by an automated program to illegally access systems and that each request is coming from an individual human, since there are many illegal and unwanted programs trying to gain the access and misuses the resources.

CAPTCHA techniques are usually classified into three main categories: Text-based, Image-based, and Sound-based [2]. Each type is suitable to serve different groups of users. Text-based CAPTCHAs are those where in the users are asked to correctly identify and enter the presented text that is displayed as an image. Figure 1.1 shows some examples of the Text-based CAPTCHAs used in some of the most popular websites like Google, Facebook, Hotmail and Paypal.

Figure 1.1 Examples of Text-based CAPTCHA

Sound-based CAPTCHAs are usually for those users who have difficulties to read as the users will have to listen to sound that is played and type in accordingly. Image-based CAPTCHAs, where the users will be asked to perform certain activity like to identify the objects in the image that is displayed (which contains a group of different images combined to form

single image). This kind of CAPTCHAs might not be suitable for those who may interpret the given Images in the CAPTCHA differently. However, different people have different capabilities to respond to different types of CAPTCHA regardless of Text-based, Image-based, or Sound-based. More in details have been elaborated in Chapter 2.

As Text-based CAPTCHAs are the easiest and simplest form of CAPTCHAs for most of the users and it is being used commonly in most of the websites; although it may not be efficient for those who have difficulty in reading. Many Text-based CHAPTCHAs have been developed and implemented using different ways like using noise, distortion and many other ways to make the presented Text-based CAPTCHA difficult to any bots and other programs to read the text as shown in Figure 1.2. However it has made it difficult for the users as well to identify the text in the presented Text-based CAPTCHA and the users will have to retype the CAPTCHA; this kind of complication can be avoided by understanding the user personal characteristics.

Figure 1.2 Examples of Text-based CAPTCHA difficult for user to understand

1.2 Problem Formulation and Motivation

Presently, CAPTCHA is an important mechanism to gain access to the required system. However, there are some difficulties for users in identifying CAPTCHA and may type it wrong and resulting to denial to access the system although they are authorize persons. Since the Text-based CAPTCHA is the most popular mechanism amongst all the CAPTCHA techniques, the difficulties of this Text-based CAPTCHA are studied and drawn out. The results of this study

have shown that the presented character(s), genders of users, their educational background and their working habits are some of the important factors determining the correctness of CAPTCHA typing by its users. Therefore, generating a Text-based CAPTCHA must use the appropriate character(s) that also combining with the educational background, genders of the users and other factors.

1.3 Objective

The main objective of this research is to determine personal characteristics that could influence the correctness of Text-based CAPTCHA verification. Taking into consideration the normal facts about human beings like their gender, age group, educational background, nationality and their visibility to determine how correctly, can the users identify the Text-based CAPTCHA so that no authorized users can have difficulties in getting access to the system.

1.4 Scope of thesis and Constraint

Since the CAPTCHA is widely used in many website and well known in identifying human from automatic program, CAPTCHA techniques are classified in several main categories, Text-based, Image-based and Sound-based (more in details about CAPTCHA and its types are elaborated in Chapter 2), in this experiment main focus is only in Text-based CAPTCHA as it is the most commonly used and most popular amongst the other two. Scope of thesis is as listed below:

1. Experimental System to generate Text-Based CAPTCHA was designed using PHP 5.2.6 and MySQL 2.5.10 to store the data.
2. Text-based CAPTCHAs are generated randomly by software without human's interfering.
3. Samples collected from three countries (Bhutan, India and Thailand), which is set to be 500 volunteers (approximately).
4. Volunteers are selected based on following categories :
    - Age groups: 10-25, 26-40, 41-60 and 60+.
    - Working environments: Student, Business, Corporate/Private, Government Officials, and House wives/Retiree or Other.
    - Whether they Work in Computer or not.

- Whether they are IT or Computer Science Professionals or not.
- Visibility: Normal, Short, Long, Short and long tilt, Short-inclined, Tilted length-Tilted.

5. The Data Collection period was for 3 months.
6. The presentation is in English for Bhutan and India while English and Thai statements are displayed for Thai volunteer.

1.5 Benefit or Expected Outcomes

This research can provide a new dimension towards the development Text-based CAPTCHA. CAPTCHA prevents unauthorized access of the system from BOTs and other unwanted programs however there are times when the authorized users face difficulties in getting access to the system due to incorrect identification of CAPTCHA. Therefore this research, it enhances developer's ability to develop Text-based CAPTCHA in better ways so that no authorized users have difficulty in getting access to the system.

1.6 Definition

CAPTCHA: CAPTCHA (Completely Automated Public Turing test to tell Computer and Humans Apart) is a program that differentiates human from computer by generating and presenting a test that only humans can solve and no computers or automated software can.

CAPTCHA Text: This means the characters (both alphabets and numbers) that are presented in the Text-based CAPTCHA.

CAPTCHA Color: It means the colors of each character of the Text-based CAPTCHA that is being generated by the proposed system. The color is generated with the combination of three primary colors Red (255, 0, 0), Green (0, 255, 0) and Blue (0, 0, 255).

Cross Tabulation Methods: A cross tabulation is also called as cross tab, in this method it displays the joint distribution of two or more variables. It is presented as a contingency table in a matrix format. Frequency distribution shows the distribution of each variable, and contingency table describes the distribution of two or more variables simultaneously. Each cell in

this method shows the number of respondents that gives a specific combination of responses, that is, each cell contains single cross tabulation.

Chi-Square: can be described as a statistical quantity equal to the summation over all variables of the quotient of the square of the difference between the observed and expected values divided by the expected values of the variable.

Chi-Square test is the most important and most used member of the no parametric family of statistical tests. Chi-square is to test the difference between an actual samples and hypothetical which may be expected because of the chance or probability. In other words Chi-square is used to test the differences between two or more actual samples.

1.7 Structure of the Thesis

The rest of Thesis is organized as follows: in Chapter 2, it provides the Fundamental Knowledge of CAPTCHA and Literature Review for the thesis. Then, Chapter 3 describes the Proposed Method followed by showing the Evaluation Results and the Analysis in Chapter 4. Finally, Discussion and Conclusions are drawn in Chapter 5.

CHAPTER II

LITERATURE REVIEW AND FUNDAMENTAL KNOWLEDGE

In this Chapter, it provides the Fundamental Knowledge and Reviews of some related works. First, Section 2.1, the general knowledge in CAPTCHA is presented. Then, the usage or Applications of CAPTCHA is stated in Section 2.2. Related work in needs of Computer Security and Authentication methods are elaborated in Section 2.3. Finally, some literature reviews on CAPTCHA are drawn in Section 2.4.

2.1. CAPTCHA

The Internet has added ease and comfort to mankind in managing many things, such as online shopping, registration, online banking, social networking, emails, etc. However, there are some automated programs that are used to access online users' data or information using illegal login mechanism(s). In order to protect this unidentified access and to distinguish human from unwanted malicious computer programs, CAPTCHA was introduced in 2000. The CAPTCHA is easy to identify for human and difficult for the computer programs for understand. As mentioned earlier, the CAPTCHA is being classified into three main categories:

1. Text-based CAPTCHA.
2. Image-based CAPTCHA.
3. Sound-based CAPTCHA.

2.1.1 Text-based CAPTCHA

The most popular technique is widely used in many web sites. It is created by randomly selecting a sequence of letters and numbers, rendering them, distorting it, adding noise, and giving some effects to the texts of the Text-based CAPTCHA, such as EZGimpy [3], Gimpy [4], Pessimal print [5], reCAPTCHA [6], etc. Consequently, those texts are difficult for any bots to understand; however, more and more use of noise, distortion, and other effects became complicated and added difficulties to the users to correctly identify the text in the CAPTCHA. Figure 2.1 shows some examples of the Text-based CAPTCHA with noise and other special effects.

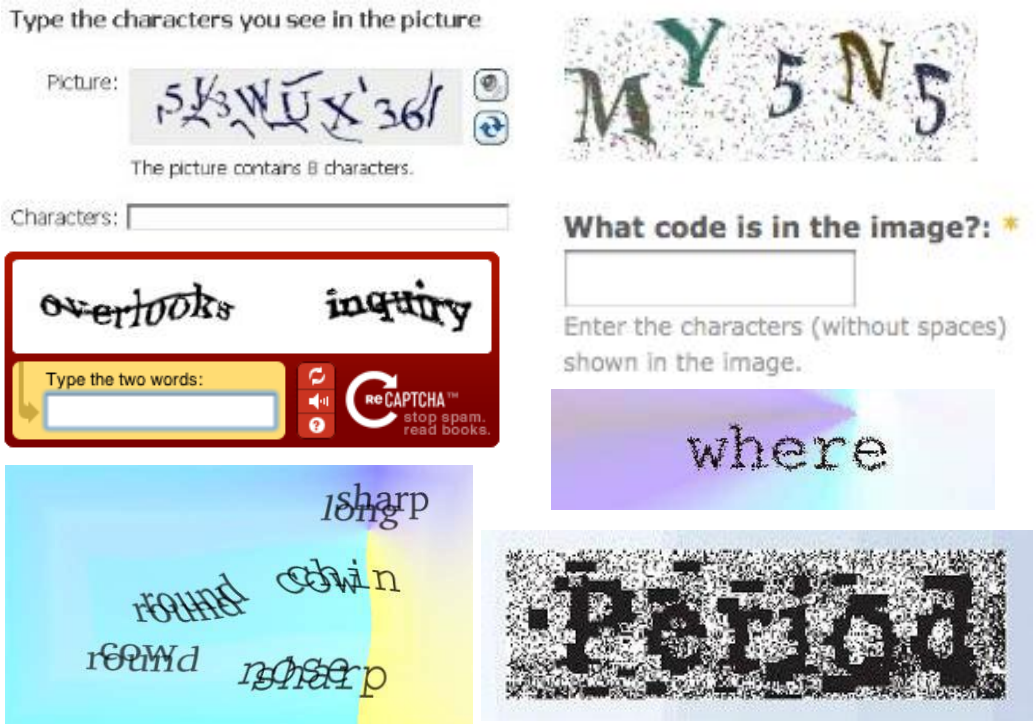


Figure 2.1 Examples of Text-based CAPTCHAs with Pessimal print [7], reCAPTCHA [1] Gimpy [8], noise [9].

Referring to Figure 2.1, one similarity among these Text-based CAPTCHA is that some texts are difficult or unable to be read or understand by normal users. They are too complicated or annoying to be solved by human. As a result, the users might ignore this CAPTCHA and wait for a new line.

2.1.2 Image-based CAPTCHA

Different from the Text-based CAPTCHA, the Image-based CAPTCHA is implemented in some interesting web sites. This type of CAPTCHA is presented using the image of some objects with a question being asked to be answered. Thus, the users must answer the question based on the presented image in the Image-based CAPTCHA as shown in Figure 2.2 below. However, the problem of answering the Image-based CAPTCHA's questions might be depended on the understanding and experience of the users. Moreover, the users may not be familiar with some of the images presented in the CAPTCHA that normally depends on the geographical features and countries. As a result, the users may wrongly answer, and may be

denied to get access to the system. Terminally, those users might get irritated and ignore. Therefore, the Image-based CAPTCHA might not be suitable for uneducated or less-educated users, including the people coming from different countries.

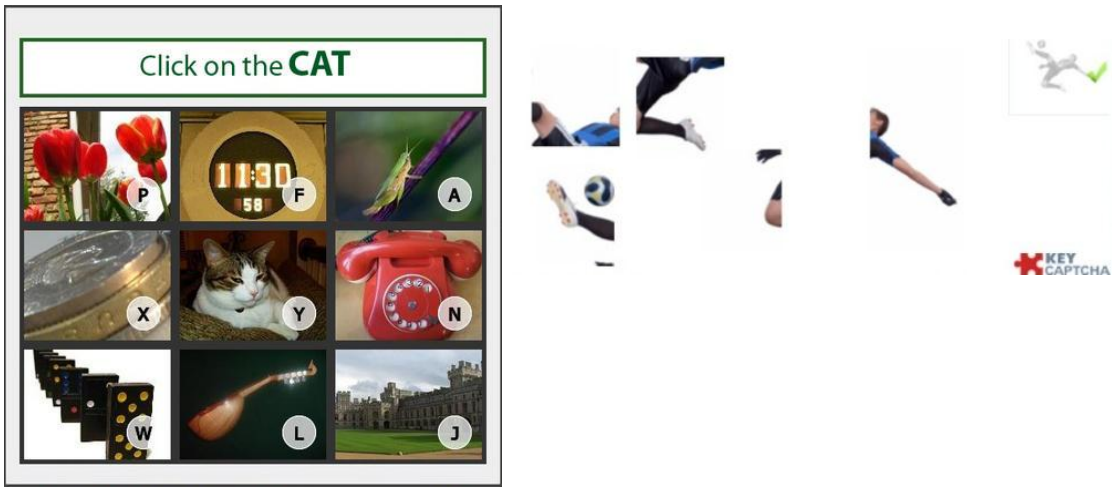


Figure 2.2 Examples of Image-based CAPTCHAs

2.1.3 Sound-based CAPTCHA

Although the first two types of the CAPTCHA are implemented and used in various most of the systems, they are not suitable for the people who are blind or have vision problems. Thus, the Sound-based CAPTCHA is proposed and implemented for this specific group of users. The implementation of the Sound-based CAPTCHA is the integration between vocal and noise. As a consequence, the difficulty in distinguishing the real value of the CAPTCHA occurs. Using this method, the users must type the correct value of the presented CAPTCHA based on the sound that they have heard.

2.2. Applications of the CAPTCHA [10]:

- Preventing Comment Spam in Blogs: Using the CAPTCHA, only humans can enter the comments in the blog.
- Protecting Website Registration: Many companies, such as Gmail, Yahoo, etc., provide free email services. However, most of these services recently suffering from the special attack, called “bots” that would sign up thousands of email

accounts in every minute. Therefore, the CAPTCHA is applied in order to solve this problem and to ensure that only humans can obtain the free accounts.

- Protecting Email Addresses from Scrapers: Spammers crawling in the web look for the email address and send a lot of spam mails that once they hold those email address. The CAPTCHA provides mechanism that will hide the email address in the website, so that the web scrapers are not able to identify it.
- Online Polls: Online poll is about asking the votes for some specific questions. There are possibilities in which some unwanted programs can push 1000s of the votes at a time. Thus, the CAPTCHA was implemented in order to protect from this kind of the attacks, so that each request is generated by humans.
- Preventing Dictionary Attacks: Its can also be used to prevent dictionary attacks in the password systems where in when the users attempt to login using wrong passwords.
- Search Engine BOTS: Sometimes many websites doesn't want to be indexed by the search engines; that is why they use CAPTCHA to ensure that no Bots enters the websites.
- Worms and Spams: The CAPTCHAs also protects against email worms and spams. The CAPTCHA systems are also used to prevent spammers from sending spams via mail servers [11] and preventing Denial of Services (DoS) attacks [12], and also to identify a child user for parental control [13].

Based on three types of the CAPTCHA presented above, none of the CAPTCHA style is suitable for every person and capability. In addition, some implemented CAPTCHAs can obstruct legal users from accessing their own system. Thus, many researches have conducted and developed new method in order to make the security level intense, protect data, and give access to the authorized users only.

2.3. Related works in security control

Increasing in use of computer to conduct many activities has become the trend in today's world. This increasing leads to the security issue to be a major challenge in the world of

Computers and Information Technology. Computer security concerns the process to prevent and detect unauthorized access of a computer, including the applications and the data. Moreover, the computer security holds the rights to give the permission for accessing to the authorized user only.

There are many methods and equipment that are introduced to maintain high level of the security. Many protection mechanisms have been widely implemented since the problems of stealing information or machine annihilation arise. Butler W. Lampson [14] said that the number of users is increasing rapidly. So, screening valid users from unauthorized users becomes a serious issue.

After realizing that the computer must be protected, various protection methods both of physical and logical are proposed.

The efficiency of unwanted software is developed as same as the development of the logical protection techniques. A research in computer security mentioned that these intrusions exist according to the monetary gains [15]. Thus, many anti-virus, data security software, and firewalls are established to prevent this unlawful situation [16].

Originally, the prevention of each computer is relied on the use of the password for logical protection and the key card for physical protection. For example, the smart card was proposed by [17] for the financial protection of electronic transaction of banks. This prevention must be performed when the users wanted to access the system or stored data.

In late 1999, *http://www.slashdot.org* released an online poll asking which was the best school in computer science; some clever people programmed an automatic ballot to answer that question without human interference. As a result, the CAPTCHA was implemented in order to avoid these kind automated programs and let only humans to answer to what is required [1].

After a long use of the CAPTCHA, the 3rd party attack problem is arising. Therefore, Chirstopher F. Tuner [18] in order to protect this kind of problem caused by human intrusion “iCAPTCHA” system was introduced. The concept of this technique is the correctness of the input sequence comparing with the CAPTCHA image that appeared on the screen.

2.4. Related works and Improvement of the CAPTCHA

Since the Internet became the major part of human's life, the information and data security also has become an important issue. However, there are some automated programs that are used to access online users' data or information using illegal login mechanism(s). In order to protect this unidentified access and to distinguish human from unwanted malicious computer programs, the CAPTCHA, which is abbreviation of "Completely Automated Public Turing test to tell Computers and Human Apart", was implemented. The CAPTCHA is easy for the humans and difficult for the computer programs.

There are many researchers developed the new module of the CAPTCHA which could be difficult for the bots to understand. However nothing could ever last as the strongest, S. Shirali-Shahreza and A. Movaghar, proposed a new CAPTCHA, "A New Anti-Spam Protocol Using CAPTCHA" [19]. The objective of this technique is to ask the user to answer the question that is easy, while the computer or the automatic spam cannot do it [20]. Fortunately, the proposed new technique seems to be working well for many years. In addition, it was widely being used in much protection system as one of the basic protection mechanism after users enter their passwords.

According to efficiency of the CAPTCHA, it gives more challenge or an opportunity for the hackers to break into this kind of protection mechanism. Therefore, various styles of CAPTCHA have been designed and developed, as mentioned previously, the CAPTCHA have been classified in three main categories: Text-based CAPTCHA, Image-based CAPTCHA, and Sound-based CAPTCHA.

Text-based CAPTCHA is the most popular style of the CAPTCHA that has ever been created by a computer program. The Text-based CAPTCHA randomly selects the letters and numbers and its sequence. It then applies some special effects like, renders them, distorts the image, and adds the noise. Since this CAPTCHA style is simple and easy approach, the authentication systems embedded the CAPTCHA are continuously increased [21][22][23]. Nevertheless, the weakness of the Text-based CAPTCHA is that the bots or the automatic programs can read the distorted letters or remove the noise using optical character recognition (OCR) [24][25]. Similar to building better protection mechanisms, the CAPTCHA-Attack algorithms are developed as well. Cui et al. had summarized that there are four CAPTCHA attack

methods. The first two methods are statistical-based method, named as the anti-noise and the anti-jamming capability. The third method is the combination of the first two approaches while the forth method is the neural network applications. With increasing of attackers' abilities, the CAPTCHA systems intend to increase the efficiencies for the CAPTCHA styles by adding more noises, distortions, and colors [26][27][28]. However, some generated CAPTCHA styles are not human friendly and easy.

The CAPTCHAs are invented in a variety of forms. The shape CAPTCHA [27] is obtained from transforming various shapes using Gestalt and Geon principles. While the spatial CAPTCHA is a text image that is derived from three-dimensional model, as shown in Figure 2.3.

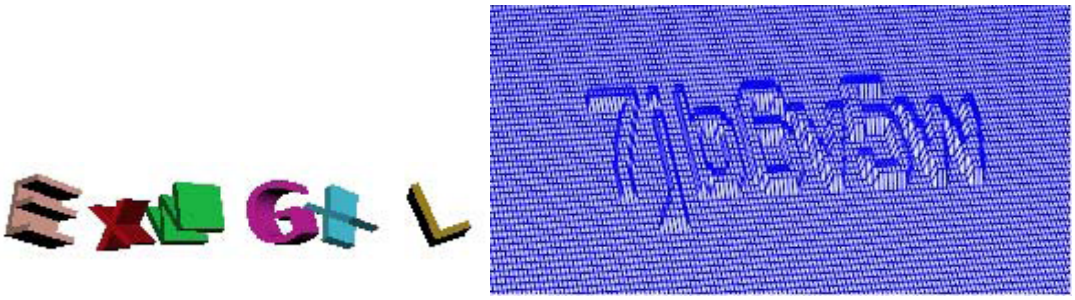


Figure 2.3 Examples of 3D CAPTCHA

In the recent days, the PDAs and the smart phones are famous. Mohammad Shirali-Shahreza [29] came up with a new concept of highlight CAPTCHA which the users have to highlight the letters appearing in the CAPTCHA.

Another one approach is Advanced Nastaliq CAPTCHA [30]. This approach uses Nastaliq font, Persian and Arabic styles, that is difficult for the OCR program to recognize. However, it becomes difficult for human to identify. In addition, many users may not be familiar to the Nastaliq font and may not be able to identify the character.

Vimina E R, Alba Urmese Areekal [31] came up with a new idea that towards the Image-based CAPTCHA. Using this technique, the users have to identify the common similar activity of the present images in the Image-based CAPTCHA as shown in the Figure 2.4. However the use of the Image-based CAPTCHA may not be always give the best results because the users may not be familiar with the images shown in the Image-based CAPTCHA. Depending on the geographical features and countries where the users belong, many pictures of the Image-

based CAPTCHA can be best described differently depending on the personal character, experience and knowledge of the users [32]. Furthermore, some pictures are unfamiliar to some users due to different geographical features and countries they belong.

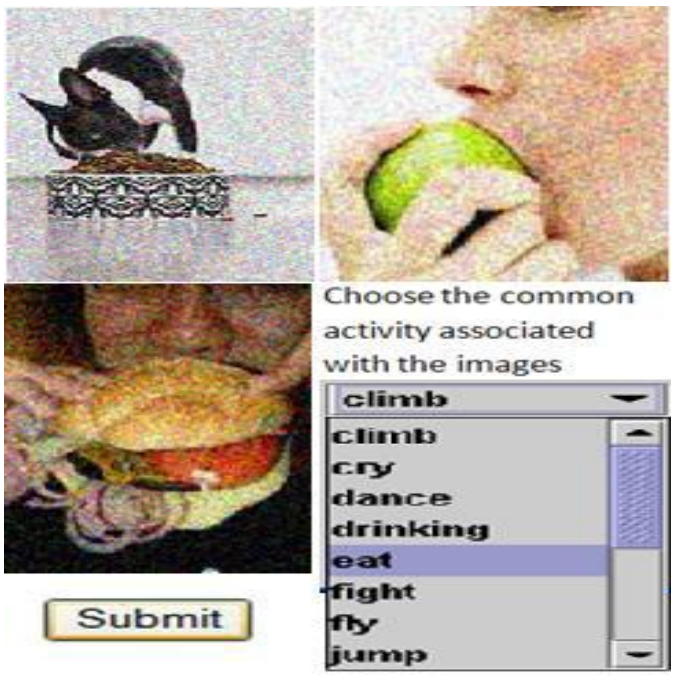


Figure 2.4 Examples of Activity CAPTCHA in Image-based CAPTCHA

Another interesting CAPTCHA style is the "speech CAPTCHA" which was implemented for the blind people. This is one of the kinds of the Sound-based CAPTCHA. This technique was proposed in 2007 and it joined the long lasting gap between the blind (visibly impaired) people and normal people [33]. The main objective of Sound-based CAPTCHA is to serve certain group of people who have problems with their vision or can be classified as blind people, so that they can also equally use a secure computing system. The system will play a sound and the blind people or the person who is using the system will have to identify the sound and type right key accordingly.

According to the problem of unidentifiable Text-based CAPTCHAs, this method is proposed to identify the impact of geographical feature on the Text-based CAPTCHA identification. Therefore, the Text-based CAPTCHA can be designed in a prospered style for all Internet users.

CHAPTER III

METHODOLOGY

This chapter describes the system environment of the proposed solution towards the Text-based CAPTCHA. It has been divided into sections were in, Section 3.1 describes the Experimental Design, Section 3.2 Design for population and samples. Section 3.3 describes Analysis of the existing Text-based CAPTCHA while Section 3.4 describes Focusing Features of Text-based CAPTCHA and Section 3.5 describes the Proposed Method. Section 3.6 describes how Data Collection has been done and lastly in 3.7 Implementation Codes of Registration forms and CAPTCHA generator is show.

The Geographical features are counting as a factor that might have some impacts on the correctness of CAPTCHA inputs. The measurement of these impacts is values of bioinformatics of users under the certain controlled environment. In this study, thus, a Text-based CAPTCHA system was implemented using PHP version 5.2.6 under web-based environment. Moreover, the Text-based CAPTCHA generating system was also developed using PHP version 5.2.6 for standalone and online system. Furthermore, all data were stored in MySQL 2.5.10 for windows.

3.1 Experimental design

As mentioned earlier in this research the main focus is only in the Text-based CAPTCHA. A system must be developed to generate the Text-based CAPTCHA. The main concept or objective of this research is to identify the factors that should be taken into consideration while developing Text-based CAPTCHA so that every time user types in Text-based CAPTCHA there is minimum number of errors and receives minimum denial to access the system. Therefore, a platform was designed to record user's details; a web based application was developed using PHP 5.2.6 to enter the users details, before entering the main Text-based CAPTCHA system. Factors that could have impacts were determined as nationalities, gender, age group, educational back ground and their working habits. Data were to be collected considering the factors as determined above.

3.2 Design for population and samples

It is important to collect data from different countries that have completely different characteristics from each other, in order to identify the impact of personal characteristic towards correct identification of the Text-based CAPTCHA. Three countries Bhutan, India and Thailand have been selected for collecting the data as all these three countries have unique characteristic from each other as described below.

3.2.1. Bhutan

Bhutan is one of the countries that is still lacking behind in technology, and the country is still developing in every aspects. Bhutan is listed in one of the smallest countries in the world with just total area of 38,394 square km with total population 742,737 (2012 estimates). People in Bhutan are most in the middle age group of 15-64 years - 65.3% (male 245,054/female 217,864), **0-14 years:** 28.9% (male 104,622/female 100,383) and **65 years and over:** 5.7% (male 21,347/female 19,157) (2011 est.) [34]. Bhutan has the literacy rate of just 47% which includes 60% male and 34% female.

The Internet was introduced in Bhutan not earlier that late 1999 as shown in the Figure 3.1. It also shows the number of the Internet users in Bhutan which has increased over the time.

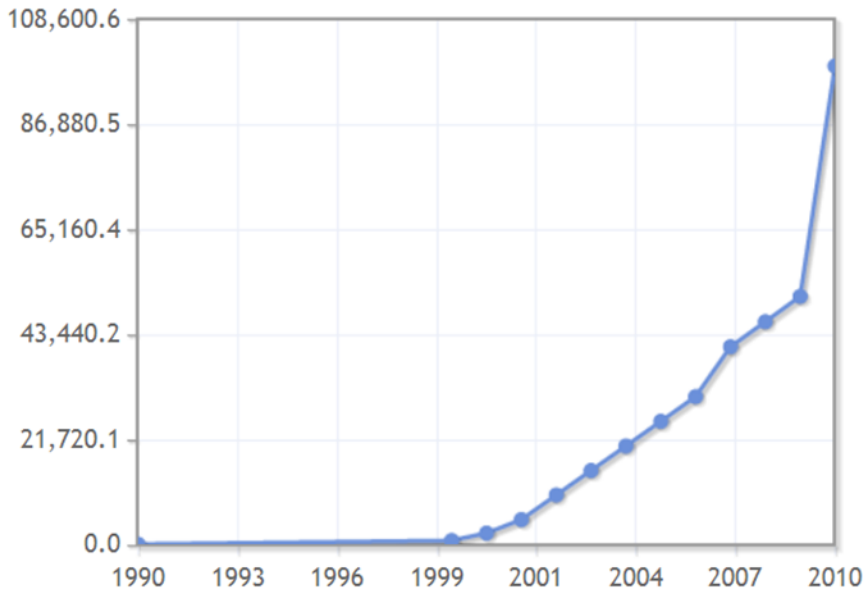


Figure 3.1 The Internet Users in Bhutan (est 2011).

3.2.2. India

India is the seventh largest country (1,269,345 sq. miles) [35] in the world with the total population more than 1 Billion (1,241,491,960) [36]. India is much more developed than Bhutan. The Internet users have been records as 78.7 million in October 2012 to 87.1 Million in December 2012 which is even expected to increase to 165 million in 2015. Literacy Rate in India as per Census 2011 is 74.04 percentage which includes 82.14 percent male and 65.46 percent female [37]

3.2.3. Thailand

Thailand is economically and technologically more developed than both Bhutan and India. The total population in Thailand was last recorded at 69.5 million people in 2011 [38]. Literacy rate for Thailand is 92.6% which includes 94.9% male and 90.5% of females (2000 Census) [39].

The Internet users in Thailand was last reported at 14653913.61 in 2010, according to a World Bank report published in 2012 [40], as shown in Figure 3.2.

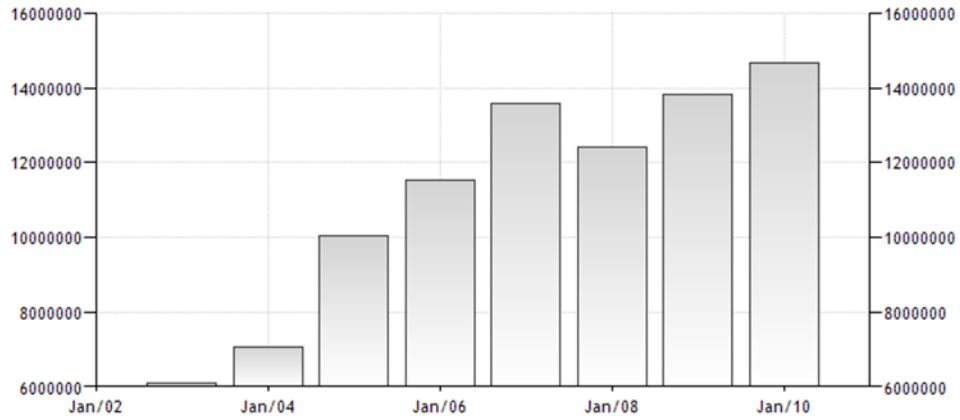
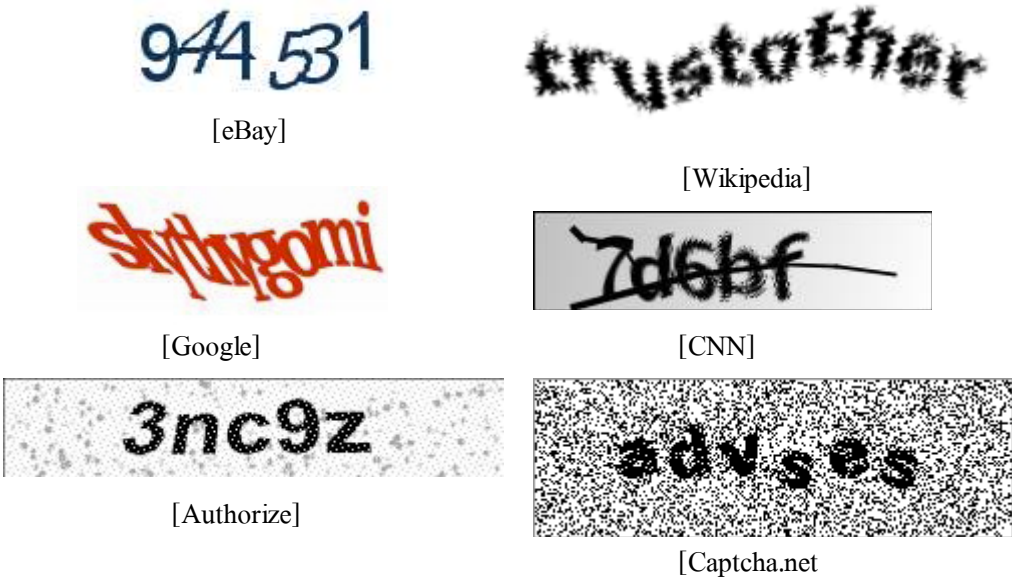


Figure 3.2 Internet Users in Thailand (est 2010).

3.3 Analysis of the existing Text-based CAPTCHA

There are many different ways of developing CAPTCHA and selecting suitable types of CAPTCHA for the website. As the Text-based CAPTCHA is one of the CAPTCHA that people are very much familiar; the Text-based CAPTCHA is most popular and commonly used in the website when comparing with other types of CAPTCHAs. In this research the main focus is only towards the Text-based CAPTCHA.

There are many types and different features added to the Text-based CAPTCHA in the recent days to add complexity for bots and other unwanted programs to understand. The Text-based CAPTCHA in the recent days have many features. Each Text-based CAPTCHAs used in different popular websites have their own unique features; like some Text-based CAPTCHA using only alphabets, some only numbers and some uses alphanumeric, many add noise like dots and curves, some applies distortion effects, and some uses the angle of tilt to the characters in the image. The Figure 3.3 below shows some of the examples of the Text-based CAPTCHA with different effects.



[eBay]

[Wikipedia]

[Google]

[CNN]

[Authorize]

[Captcha.net]

Figure 3.3 Sample of the Text-based CAPTCHA format used in many websites.

3.4 Focusing features of Text-based CAPTCHA

Although there are various styles of CAPTCHA over the real world, in this experiment will focus only the simple Text-based CAPTCHA schema without the rendering mechanism. As the main concept and the objective of the research is to identify the factors that could have impact on the correct identification of the Text-based CAPTCHA the generation of the Text-based CAPTCHA is kept very simple so that the identified factors have impact or not can be easily measured. Alphabets, a-z, and simple digits, using only number 0-9, use of color to text and fixed white background color and with no distortion and no noise has been given however just to maintain the complexity the angle of tilt up to 10 degree has been set to randomly selected character of the Text-based CAPTCHA.

3.5 Proposed Method

The main concept of this thesis is to identify factors towards the correct identification of the Text-based CAPTCHA. Hence the Text-based CAPTCHA testing method has been divided into two parts as it is important to collect CAPTCHA inputs as well as the general demographic information about the users who have been selected to type the Text-based CAPTCHA in the designed system. Factors that are assumed to have impacts towards the correct identification of the Text-based CAPTCHA are as follows:

1. Nationality
2. Gender
3. Age group
4. Career or Educational background
5. Working habits of the users whether they user computer to conduct their daily works
6. Computer Science/IT personal or others
7. Vision capability
8. Color Blindness
9. Colors of the Text-based CAPTCHA
10. Characters presented in the Text-based CAPTCHA

Each factor have been selected and assumed to have significant impact in the correct identification of the Text-based CAPTCHA because there are many users in the world who have different characteristics and have different abilities to identify CAPTHCAs. So it is important to categorize users by Nationality as the environment and development of the countries determines people's ability towards anything. It is universal fact that different people in different age groups have different ability. Many people use the Internet and come across CAPTCHAs; all the people may not always have same educational background. So, it's also important to know how people from different working environments and their working habits identify the CAPTCHA. It has been observed that different people have different vision abilities. Thus, there could be possibility that people from different vision abilities could have different capabilities to identify CAPTCHA.

3.5.1 Demography information

First part of the data collection starts with the registration of the volunteers. The registration procedure is compulsory for all volunteers as it determines characteristics of each volunteer. The volunteers must provide the following information as listed below.

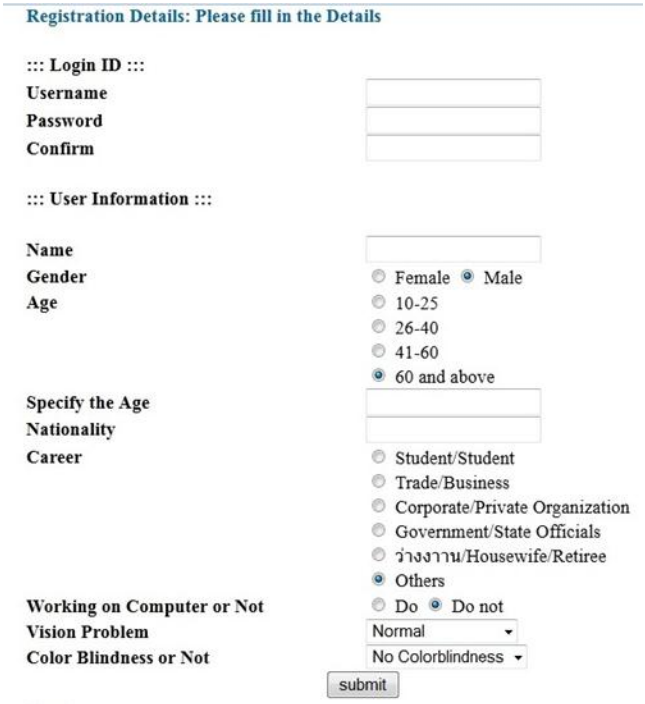


Figure 3.4 Registration Form for user enters their information.

Therefore, the volunteer must provide his/her demographic information in the first part of the questionnaire as follows: user name, password, gender, age, age group, nationality, career, vision, colorblindness, works in computer, vision problem and colorblindness as in Figure 3.4 the registration form presented for each volunteer.

3.5.2 Text-based CAPTCHA generator

In this experiment the Text-based CAPTCHA generator have been divided into two parts: CAPTCHA text, and CAPTCHA colors.

3.5.2.1. CAPTCHA Text

The CAPTCHA presentation must always be easy to read and understand by human but difficult for malicious software. In this experiment, five CAPTCHA characters with

three alphabets a - z and two numbers 0 – 9 are randomly selected. Positions of alphabets and numbers are also random as well. To keep the recognition of the CAPTCHA simple for human, only one type of font with no style (bold, italic or underline) and constant font of San.ttf is used. However, the CAPTCHA characters which are randomly selected are tilted to 10 degrees to add complexity so malicious programs cannot recognize. Nevertheless, this generated Text-based CAPTCHA is easily identified by human even with the tilt. Figure 3.5 below shows the sample of the Text-based CAPTCHA generated by the proposed system.

Figure 3.5 Example of CAPTCHA generated by the proposed system.

3.5.2.2. Color Categories

Another important attribute of the Text-based CAPTCHA is color. One color is usually obtained from the combination of three main primary colors: Red, Blue, and Green, as called RGB. Thus, there are many different colors to be generated. Generally, the presented Text-based CAPTCHA will consist of two types of color: the font color, and the background color. In this experiment, the background color of CAPTCHA is set to one fixed color – White (255,255,255). Colors used for the Text-based CAPTCHA have unique variations as shown in Table 3.1. This research has ranged and grouped the colors as shown in the Table 3.2.

Basic fundamental used to categorize colors to each CAPTCHA text is that no text in the same Text-based CAPTCHA has the same color. As the background color is set to be

fixed white no CHATCHA text is generated with the color white as well. Figure 3.5 shows samples of generated Text-based CAPTCHA based on the proposed method.

Table 3.1 All Colors and Codes

Color Name	Dec Code	% Code	Color
Black	RGB (0, 0, 0)	RGB (0%, 0%, 0%)	
Grey	RGB (128, 128, 128)	RGB (50%, 50%, 50%)	
Silver	RGB (192, 192, 192)	RGB (75%, 75%, 75%)	
Maroon	RGB (128, 0, 0)	RGB (50%, 0%, 0%)	
Red	RGB (255, 0, 0)	RGB (100%, 0%, 0%)	
Olive	RGB (128, 128, 0)	RGB (50%, 50%, 0%)	
Yellow	RGB (255, 255, 0)	RGB (100%, 100%, 0%)	
Green	RGB (0, 128, 0)	RGB (0%, 50%, 0%)	
Lime	RGB (0, 255, 0)	RGB (0%, 100%, 0%)	
Teal	RGB (0, 128, 128)	RGB (0%, 50%, 50%)	
Aqua	RGB (0, 255, 255)	RGB (0%, 100%, 100%)	
Navy	RGB (0, 0, 128)	RGB (0%, 0%, 50%)	
Blue	RGB (0, 0, 255)	RGB (0%, 0%, 100%)	
Purple	RGB (128, 0, 128)	RGB (50%, 0%, 50%)	
Fuchsia	RGB (255, 0, 255)	RGB (100%, 0%, 100%)	

Table 3.2 Main Primary Colors and Codes

Color	Code
RED	255,0,0
GREEN	0,255,0
BLUE	0,0,255

3.6 Data Gathering

As the main objective of this experiment is to identify impact factors related to the geographical features over Text-based CAPTCHA, varieties of information are collected such

as nationality, gender, age, typing character in each position, etc. The collected data also include entered character in each position of each volunteer. The nationalities of volunteers in the research are Bhutanese, Indian and Thai.

The data collection duration was 3 months in total which was divided as April-May 2012 to collect data samples from Bhutan and India. Thai samples were collected in December 2012 for a month. Total of 500 people were randomly selected from three different countries Bhutan (200), India (100) and Thailand (100). Each volunteer must enter 30 different-random CAPTCHAs generated by the Texted-based CAPTCHA generating system.

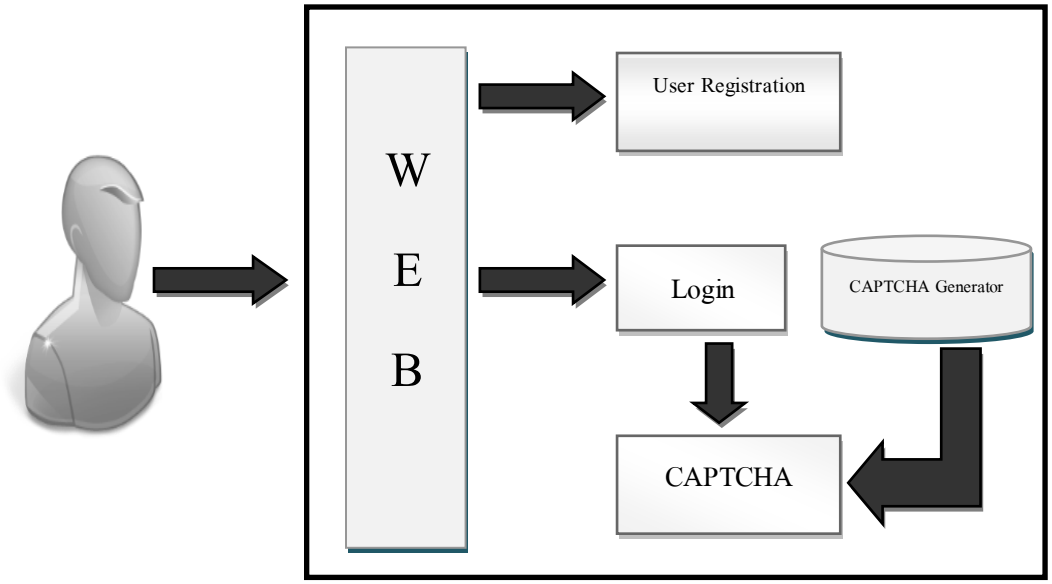


Figure 3.6 Overview of Experimental System

The system was designed as web based application and hosted over the internet so that the users could use the system even from a distance location using internet as shown in Figure 3.6. Referring to Figure 3.6, firstly, the user needs to register to the system providing few demographic information of the user as discussed earlier Section 3.4. After the user completes the registration process; user logs in with the user name and password that was provided during registration, finally the user is then preceded to the Text-based CAPTCHA testing procedure.

This login process will directly run the CAPTCHA system that generates a Text-based CAPTCHA; as discussed in the earlier Section 3.4.2. Under the testing procedure, each user has to enter 30 different Text-based CAPTCHAs. These entered data are stored in the MySQL database. These data were analyzed using the SPSS software provided by Chulalongkorn University, which is discussed in detail in the next Chapter.

3.7 Implementation Code:

All system was implemented and coded in PHP. System is divided into two main parts Registration Form and Text-based CAPTCHA generation. All the data input by the users using the designed system is saved in the database (MySQL 5.0.5).

3.7.1. Text-based CAPTCHA generation Code:

The codes below show the function to generate the Text-based CAPTCHA by the proposed system.

```
$LinearBackgroud = 0;
$GranularNoise = 0;
$LinearNoise = 0;
$SquareNoise = 0;
$distortion_hor = 0;
$distortion_ver = 0;
$CharacterColorMode = 1;$subset = 0;
$CharacterSizeMode = 1;$CharacterRotationMode = 10;
$font = 'sans.ttf';$randomfont = 0;$Codelength = 5;
$Characters = "abcdefghijklmnopqrstuvwxyz";
$Numerics = "0123456789";
$height = 80;
$BackgroudColorMode = 0;
$fontsize = $height * 0.6;
// Create some colors
$white = imagecolorallocate($im, 255, 255, 255);
$grey = imagecolorallocate($im, 238, 238, 238);
$black = imagecolorallocate($im, 0, 0, 0);
$red = imagecolorallocate($im, 255, 0, 0);
// colors to fade
$red_start   = mt_rand(0,255);
$red_end     = mt_rand($red_start,255);
$green_start = mt_rand(0,255);
$green_end   = mt_rand($green_start,255);
$blue_start  = mt_rand(0,255);
$blue_end    = mt_rand($blue_start,255);
function dif ($start,$end)
{
        if ($start >= $end)
                $dif = $start - $end;
        else
                $dif = $end - $start;
        return $dif;
}
function draw($start,$end,$pos,$step_width)
{
        if ($start > $end)
                $color = $start - $step_width * $pos;
        else
                $color = $start + $step_width * $pos;
        return $color;
}
```

3.7.1. **Registration Form Code**: The code below is for the Registration form to gather the required information of the users. All the information is being saved in the database.

```
::: User Information :::
<td>
<input name="gender" type="radio" id="radio" value="f" checked="checked"> Female

<input name="gender" type="radio" id="radio" value="m" checked="checked"> Male
</td><td>
<input name="age" type="radio" id="age" value="10-25" checked="checked"> 10-25
<input name="age" type="radio" id="age" value="26-40" checked="checked"> 26-40
<input name="age" type="radio" id="age" value="41-60" checked="checked"> 41-60
<input name="age" type="radio" id="age" value="other" checked="checked"> 60 and
above
</td>
<td><input type="text" name="nationality" id="nationality"></td>
<td>
<input name="job" type="radio" id="radio" value="s" checked="checked">
Student/Student
<input name="job" type="radio" id="radio" value="f" checked="checked">
Trade/Business
<input name="job" type="radio" id="radio" value="c" checked="checked">
Corporate/Private Organization
<input name="job" type="radio" id="radio" value="d" checked="checked">
Government/State Officials
<input name="job" type="radio" id="radio" value="h" checked="checked"> ว่างงาน/
Housewife/Retiree
<input name="job" type="radio" id="radio" value="o" checked="checked">
Others</td>
<th scope="row"> <div align="left">Working on Computer or Not</div></th>
<td> <input name="computer" type="radio" id="computer" value="do"
checked="checked" onClick="showCharacter(true);"> Do
<input name="computer" type="radio" id="computer" value="not" checked="checked"
onClick="showCharacter(false);"> Do not</td>
<td><select name="problem_eye">
<option value="nom" selected> Normal </option>
<option value="sht"> Short</option>
<option value="lon"> Long </option>
<option value="ang"> Tilted </option>
<option value="san"> Short-inclined</option>
<option value="lan"> length-Tilted </option>
<option value="sla"> Short and long tilt </option></select></td>
<td><select name="color_blindness">
<option value="non"> No Colorblindness </option>
<option value="red"> Red-Green</option>
<option value="blu"> Blue-Yellow </option>
<option value="not"> Others </option></select>
</td>
```

CHAPTER IV

EXPERIMENTAL RESULTS

This chapter describes the analysis and the results. It has been divided into sections where in Section 4.1 Analysis of Demographic Factors towards correct identification of Text-based CAPTCHA, Section 4.2 CAPTCHA analysis and Section 4.3 states the Final Conclusion.

Before analyzing the CAPTCHA characters lets first look into the demographic information of people selected for the experiment. Total of 500 data samples were to be collected from three different countries; however as shown in the Table 4.1, 4.2, 4.3, 4.4, and 4.5:, the exact figures (387) of the sample data were collected; after many data samples being rejected due to some failure in some criteria. So only those data were not selected to be analyzed:

- 238 volunteers from Bhutan, 77 from India and 72 from Thailand
- There were total of 232 Males and 155 Females.
- Volunteer were from different age groups: 80 in 15-20, 275 in 26-40, 30 in 41-60 and 2 in 60+.
- Most of the volunteers were from Corporate/Private Organization category with 198 people, government with 82 people, students with 45, trade and business with 33 people and unemployment or others were 8 people.
- More than half of the total volunteer collected uses computer in their daily working habits with 205.
- None of volunteer had color blindness problem.

Table 4.1 Gender based Categorization.

Gender	Bhutanese	Indian	Thai	Total
male	133	58	41	232
Female	105	19	31	155
Total	238	77	72	387

Table 4.2 Age group based Categorization.

Age Group	Bhutanese	Indian	Thai	Total
15-20	49	1	30	80
26-40	170	68	37	275
41-60	19	7	4	30
60+	0	1	1	2
Total	238	77	72	387

Table 4.3 Career based Categorization.

Career	Bhutanese	Indian	Thai
Corporate/Private Organization	136	53	10
Government/State Official	68	3	11
Trade/Business	8	18	7
Others	13	2	11
Student	12	2	31
House wife/Retiree	1	0	2

Table 4.4 Work with Computers based Categorization.

Work with Computer	Bhutanese	Indian	Thai	Total
Yes	153	9	43	205
No	85	68	29	182
Total	238	77	72	387

Table 4.5 Visibility based Categorization.

Visibility	Bhutanese	Indian	Thai	Total
Normal	185	59	60	304
Not Normal	53	18	12	83
Total	238	77	72	387

4.1. Analysis of Demographic Factors towards correct identification of Text-based CAPTCHA.

Chi-square Cross Tabulation method was used to identify personal characteristic of user towards the correct identification of the Text-based CAPTCHA in different positions of characters in the presented Text-based CAPTCHA. The statistical analysis uses the significant level equals to 0.05 ($\alpha = 0.05$) in each case.

4.1.1. Nationality

The users from different countries do have significant impacts on the correct identification of the presented Text-based CAPTCHA with Chi-square result which is less than α =0.05. Bhutan has the highest percent of incorrect Text-based CAPTCHA identified when compared with the users from India and Thailand in all the positions of the characters of the presented Text-based CAPTCHA as shown in the Table A1 for 1st position, Table A2 for 2nd position, Table A3 for 3rd Position, Table A4 for 4th position and Table A5 for 5th position.

4.1.2. Age Group

Different people from different age groups have different abilities to identify the CAPTCHA. The results in this research have proven that ability of people from different ages do have impacts towards the correct identification of the Text-based CAPTCHA with the Chi-square analysis result less than α =0.05. It has also been noted that volunteers in the age group of 41-60 and 60+ have more likely tendency to identify the presented Text-based CAPTCHA wrongly than the volunteers from other age groups in every position of the presented Text-based CAPTCHA as shown in the Table A6 for 1st position, Table A7 for 2nd Position, Table A8 for 3rd Position, Table A9 for 4th position and Table A10 for 5th position.

4.1.3. Career (Educational background of the volunteer)

Career or the educational background of the user is another important factor towards the correct identification of the Text-based CAPTCHA as it determines in what environment the user uses the IT equipment or the infrastructure. Therefore, volunteers in this experiment have been selected from different working environments like students, corporate or

private organization, government and state officials, trade and business people, retiree and house wife, and other which includes any other profession that are not included in the list.

The result of the Chi-square cross tabulation analysis shows that there is significant impact of career of the users in correct identification when compared in all the position in the presented Text-based CAPTCHA with the p-value 0.00 which is less 0.05. It has also shown that the users who work in the government or state officials are more likely to identify the presented Text-based CAPTCHA wrongly than any other groups of users as shown in the Figure A11 for the 1st position, Figure A12 for 2nd position, Figure A13 for 3rd position, Figure A14 for 4th position and Figure A15 for the 5th position in the presented Text-based CAPTCHA.

However, it has also been noted that the users working in government/state officials who have wrongly identified the presented Text-based CAPTCHA are mostly from the age-group 41-60 and who doesn't use computer to carry out their daily works. As shown in the Tables A26 for 1st position, Tables A27 2nd position, Tables A28 for 3rd position, Tables A29 for 4th position, and Tables A30 for 5th position, that government/state officials in the age group of 41-60 have identified the presented Text-based CAPTCHA more wrongly than the other age group users in every position. In addition to that a government/state official who doesn't work with computers daily to conduct their routine works have the highest error/mistyped the presented Text-based CAPTCHA as shown in the Tables A31 for 1st position, Table A32 for 2nd position, Table A33 for 3rd position, Table A34 for 4th position and Table A35 for 5th position.

4.1.4. Working in computer

After having identified whether career of users has impact on the correctness of CAPTCHA, it is important to further categorize it to working habits of the users**.** The result of this analysis shows that the use of computer in people's everyday activities have significant impact with Chi-square result of less than α =0.05. It has also shown that the users who uses computer in their everyday activities do have more accuracy in identifying the presented Text-based CAPTCHA than those who don't use computers in their working environments in all position of the presented Text-based CAPTCHA as shown in the Table A16 for 1st position of the presented Text-based CAPTCHA, Table A17 for 2nd position of the presented Text-based CAPTCHA, Table A18 for 3rd position of the presented Text-based CAPTCHA, Table A19 for 4th

position of the presented Text-based CAPTCHA, Table A20 for 5th position of the presented Text-based CAPTCHA.

4.1.5. IT professionals or not

After the segregation of the users into different groups according to their Educational Background and their Working Habits on computers; it is important to know whether the users are Computer Science/IT professionals or not. However, the results of the analysis shows that there is no any significant difference or impact on the presented Text-based CAPTCHA identification by the users regardless of the IT profession skill.

4.1.6. Visibility

Different people have different visibilities; how can each type of visibility of people have any impacts towards the identification of the Text-based CAPTCHA. Visibilities of people are classified as follows: normal, short, long, short and long tilt and short inclined. The result shows that there is significant impact of visibility of users towards the correct identification of the presented Text-based CAPTCHA with Chi-square analysis result which is less than α= 0.05. The cross tabulation comparison also shows that the users with the short and short/long tilted vision have the maximum impact on the correct identification of the Text-based CAPTCHA in every position of the presented Text-based CAPTCHA as shown in the Table A21 for 1st position, Table A22 for 2nd position, Table A23 for 3rd position, Table A24 for 4th position and Table A25 for 5th position.

4.2. CAPTCHA Analysis

This analysis has been broken down into two parts: color analysis, and CAPTCHA text analysis. The analyzing method is the Chi-Square test with 95% confident interval (α = 0.05). The test here shows the correctness in identifying the present Text-based CAPTCHA in each position which is compared with the real Text-based CAPTCHA value.

4.2.1. Color Analysis

As mentioned in the earlier chapter, colors are categorized into 16 different colors which are extracted from the combination of three primary colors (RGB). Analyzing the correctness of the present Text-based CAPTCHA with the colors of the presented Text-based CAPTCHA; seems like colors do not have much significant impact on the correct identification of the Text-based CAPTCHA. As shown in the Table 4.6 no matter what color is being used, the correctness to identify the Text-based CAPTCHA has same impact on all type of colors.

Table 4.6 Color by Correctness of Text-based CAPTCHA identification

		Correctness of Character			**Percentage of Correct Identification of Text-based CAPTCHA**	
		Correct	**Not Correct**	**Total**	**Correct**	**Not Correct**
Color of Character	Black	2914	183	3097	94%	6%
	Grey	7875	510	8385	94%	6%
	Silver	3744	215	3959	95%	5%
	Maroon	1209	66	1275	95%	5%
	Red	3502	220	3722	94%	6%
	Olive	3671	237	3908	94%	6%
	Yellow	5162	292	5454	95%	5%
	Green	3895	176	4071	96%	4%
	Lime	1565	70	1635	96%	4%
	Teal	4740	275	5015	95%	5%
	Aqua	4357	230	4587	95%	5%
	Navy	1772	93	1865	95%	5%
	Blue	3203	186	3389	95%	5%
	Purple	2987	192	3179	94%	6%
	Fuchsia	3742	258	4000	94%	6%
Total		54338	3203	57541	94%	6%

When tested with Chi-Square test with 95% confident interval (α = 0.05), the result is more than α=0.05 for colors of characters in every position of the presented Text-based CAPTCHA. Table 4.8 shows the percentage of correct and not correct identification of Text-based CATPCHA in each color. Which can be stated as, regardless of what color the CAPTCHA text is, there isn't much impact towards the correct identification of the Text-based CAPTCHA. Most of the users don't have difficulty in identifying the presented Text-based CAPTCHA correctly.

4.2.2. CAPTCHA Text Analysis

The CAPTCHA Text refers to the characters (alphabets and number) in the presented Text-based CAPTCHA. The result of this analysis states that correct identification of Text-based CAPTCHA is related to the CAPTCHA Text with significant level equal to zero < α = 0.05. There are some characters that are identified incorrectly by many users. Each CAPTCHA text has been analyzed in each position in the presented Text-based CAPTCHA. There are some English alphabets and numbers which are identified wrongly by the users. As shown in the Figure 4.1, 4.2, 4.3, 4.4 4.5 below, character "j" is identified as most wrongly typed character by the users in all position.

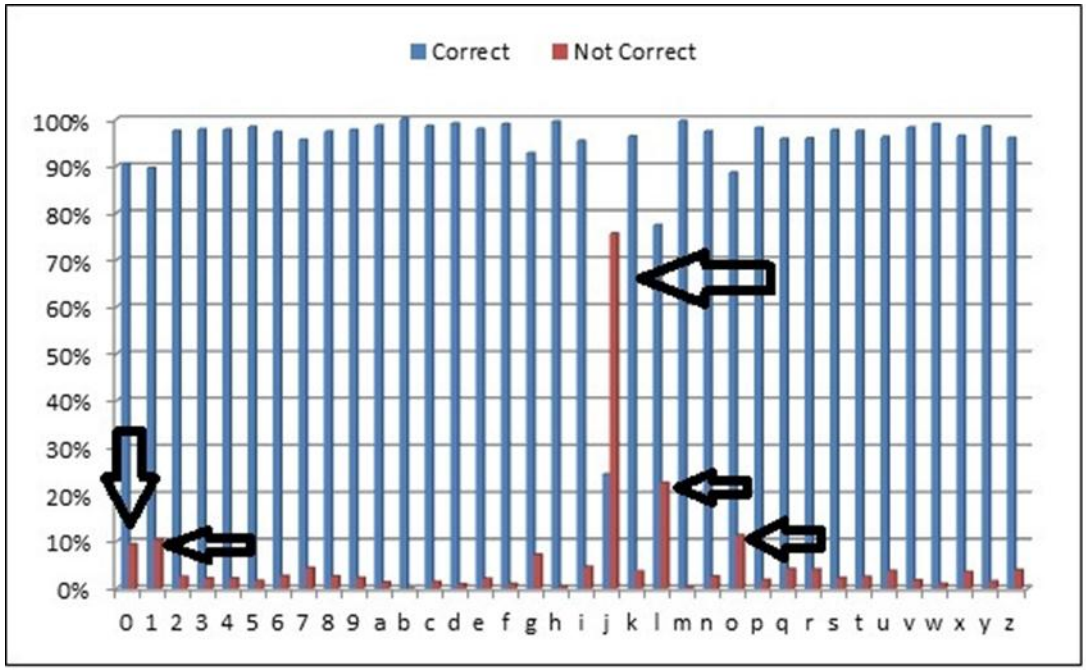


Figure 4.1 CAPTCHA Text 1 by correctness of Text-based CAPTCHA identification

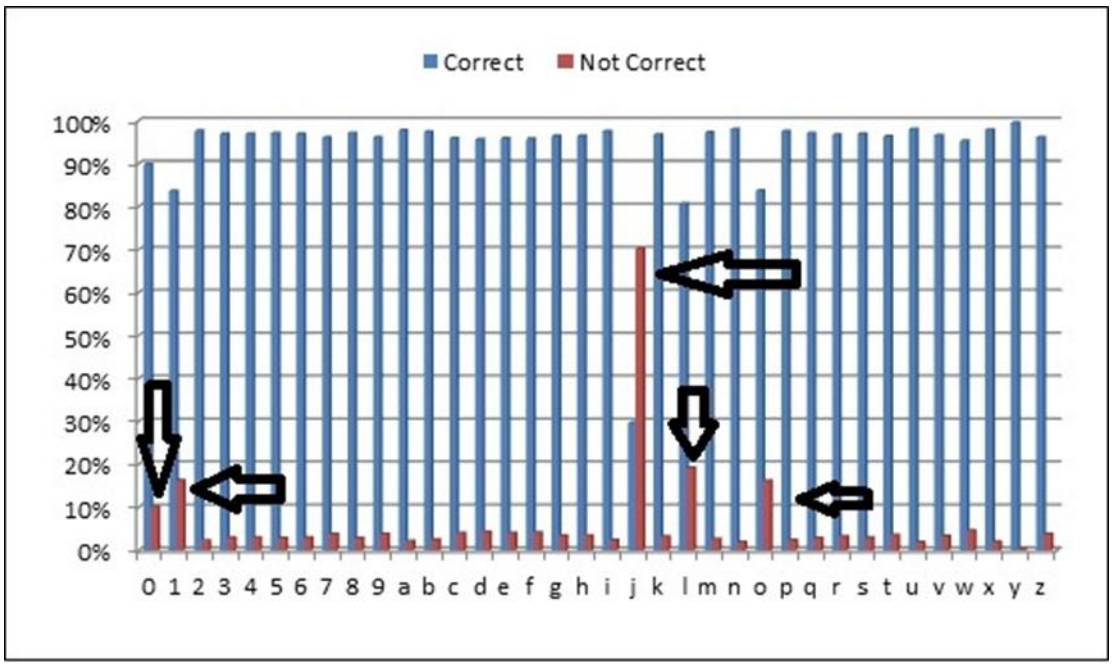


Figure 4.2 CAPTCHA Text 2 by correctness of Text-based CAPTCHA identification

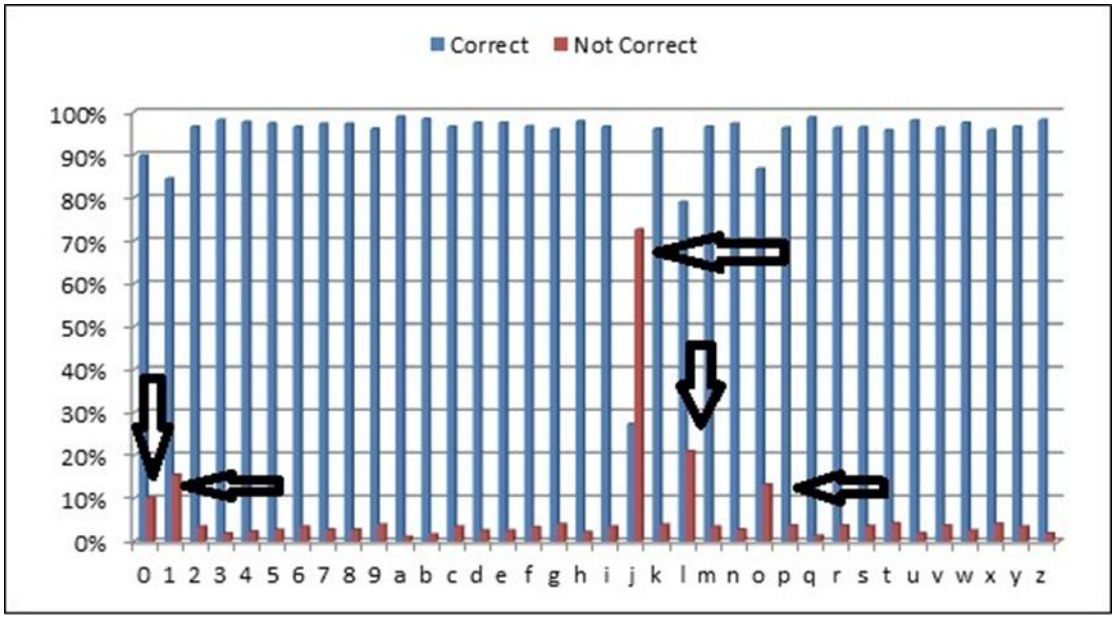


Figure 4.3 CAPTCHA Text 3 by correctness of Text-based CAPTCHA identification

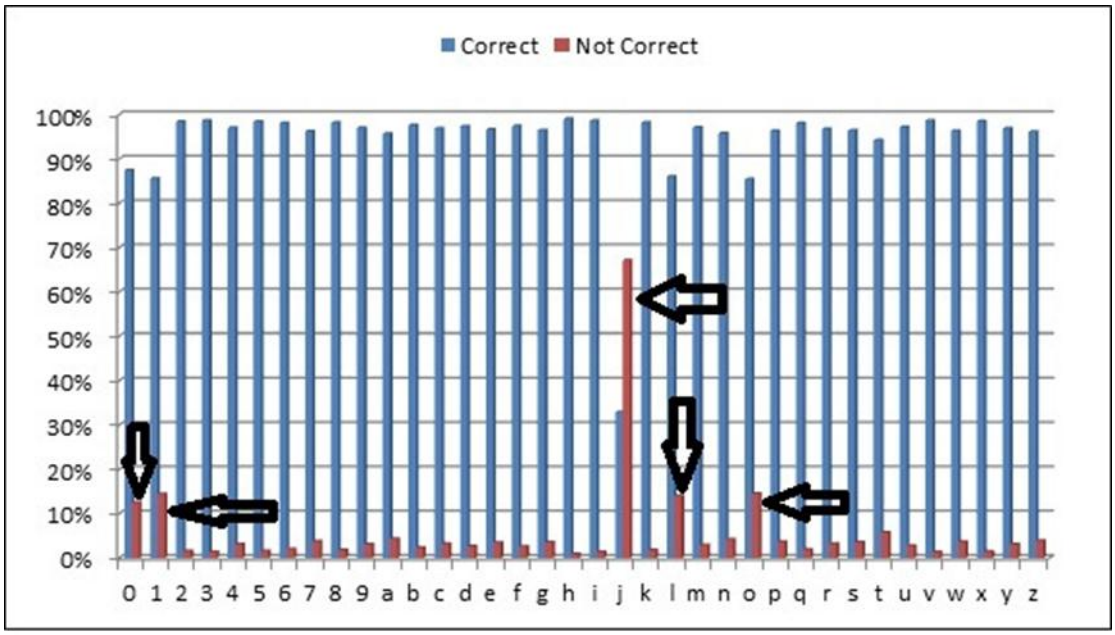


Figure 4.4 CAPTCHA Text 4 by correctness of Text-based CAPTCHA identification

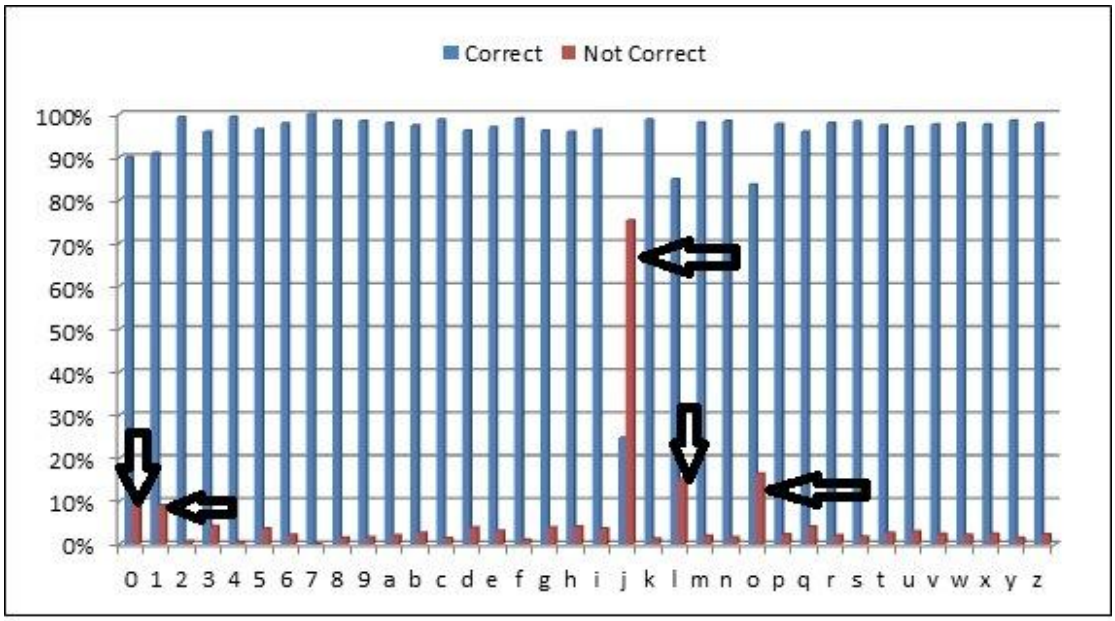


Figure 4.5 CAPTCHA Text 3 by correctness of Text-based CAPTCHA identification

Apart from the character “j” it has been noted that even the character “l”, “o”, “0” and “1” do have impact on correct identification of the Text-based CAPTCHA. These characters are identified wrongly by many users constantly in every position. Character “j” is mistyped as “i”, “o” as “0”, vies versa and character “1” with “l” and vies versa.

4.3. Final Conclusion

The results from the experiment discovers that the personal characteristics of the users like Nationality, Age-group, Career, Working Habit in computer and Visibility do have unified influences in the correct identification of the Text-based CAPTCHA. It has also noted that use of certain characters (alphabets and numbers) in the Text-based CAPTCHA could lead to more incorrect identification of the Text-based CAPTCHA. The characters in the Text-based CAPTCHA (alphabets and numbers) identified to have more impact in the correct identification are ‘j’, ‘l’, 0, ‘o’, ‘1’.

CHAPTER V

DISCUSSTION AND CONCLUSION

In this Chapter, the Discussion will be discussed in Section 5.1, Limitation of the experiment is stated in Section 5.2 and finally Conclusion will be drawn in Section 5.3.

5.1. Discussion

Presently, people have started work over internet and also browse the internet for various activities. The number of times people around the world visit internet each day has tremendously increased. The exchange of information and storage of data is one of the most important factors in the Internet. Unfortunately, there lies nothing that could be fully secured over the internet. Security level could be maintained up to certain level; however, it can't be 100% safe in any cases.

There are many methods of securities as discussed earlier in Chapter 3; one simple mechanism is the use of login password, but this method was broken by a simple malware. Thus in addition to use of login password another mechanism, CAPTCHA, has been implemented and being used till date. There are various types of CAPTCHA each type serving its own purpose. Many research and developments are carried out in CAPTCHA by many researchers; however there is no proof that could say that CAPTCHA system could completely secure the access as there are many new techniques that's been developed by people to read the images like OCR based Image processing etc.

There are many research and development in the process of making the image of the Text-based CAPTCHA difficult for identification by any bots; however the personal characteristics of the users using the CAPTCHA has never been taken into consideration when developing the Text-based CAPTCHA by the developers. This could be one of the reasons that people tend to get denial to access the system even though he/she is an authorized users. It has also been noticed that sometime when the users type the CAPTCHA they tend to get irritated to type again and again due to failure in typing the CAPTCHA correctly, this could also be because CAPTCHA development and implementation not being matched with the user's personal characteristics. Although some research claims that people are familiar in using CAPTCHA for

authentication process, there is no prove that the CAPTCHA system is really suitable for every human' capability.

Hence this research has shown the results that could determine and identify the significant influencer on Text-based CAPTCHA verification from personal characteristic of users. However, there are some limitations in this research that has been stated in the next section.

5.2. Limitation of the study

There are many styles of the Text-based CAPTCHA, how can each of this style be measured? In this experiment no noise and neither any distortion been used to keep the Text-based CAPTCHA simple.

5.3. Conclusion

It is the fact that CAPTCHA, in particular the Text-based CAPTCHA, is widely used over the internet to protect unauthorized software in gaining access to the system. However, the presentation of CAPTCHA in various aspects may not support users when they want to enter into the system. Therefore, this research focuses in the possibility that users may mistype the presented Text-based CAPTCHA due to the some hidden factors: the presented Text-based CAPTCHA character(s), and Career or Educational background of the users, Working Habits of the users whether they use computer to carry out their daily works and the visibility capability of the users.

The results from this study indicate that people with different opportunities to access the computer technology affects the correctness of Text-based CAPTCHA typing. Also, the use of character "j" may not be suitable in the Text-based CAPTCHA since it was found that most people cannot read it right. In addition to that, characters "l", "o", "0" and 1 are also identified as unsuitable to be used in the Text-based CAPTCHA. Thus, using Text-based CAPTCHA must consider choosing the right character to be presented so that the authorized users will be able to access the system when needed. Furthermore, the study has shown that the ratio of three main primary colors, red, green, blue, and the visual capability of the user in each position cause no effect in correct Text-based CAPTCHA identification for users.

Thus, the factors to be considered when implementing the complete Text-based CAPTCHA system to protect from intrusion from unwanted programs and provide facility in user's typing are characters that are used in the Text-based CAPTCHA, gender, age group, career or educational background of the users and working habits whether the users use computer to carry out their day to day works.

REFERENCES

[1] Manuel, B. Luis, V.A. and John, L. The CAPTCHA Project School of Computer Science, Carnegie-Mellon University, November 2000. Available from: The Official CAPTCHA Site: http://www.CAPTCHA.net [2012,September]

[2] Jeff, Y. and Ahmad, E.A. Usability of CAPTCHAs Or uasbility issues in CAPTCH design, Proceeding SOUPS '08 Proceedings of the 4th symposium on Usable privacy and security, July 2008.

[3] Mohammad, S.S. and Sajad, S.S. Multilingual Highlighting CAPTCHA, 8th International Conference on Information Technology: New Generations, 2011.

[4] Mohammad, S.S. and Sajad, S.S. Dynamic CAPTCHA, ICCSE internation Conference pp.676-679. 2012.

[5] CAPTCHA image, blogspot, [online]
Available from: http:// 1.bp.blogspot.com/_sMkPD1yZh7M/TS8BL72sN7I/ AAAAAAAAAAw/25wtECVR9wo/s1600/6.jpg. [March 2013].

[6] Manuel, B. Luis, V.A. and John, L. The CAPTCHA Project School of Computer Science, Carnegie-Mellon University, November 2000. Available from: The Official CAPTCHA Site: http://www.CAPTCHA.net [2012,September]

[7] Vimina, E. R., Alba, U. A. Telling Computers and Humans Apart Automatically Using Activity Recognition, IEEE International Conference on Digital Object Identifier, pp.4906-4909. 2009.

[8] Yassin, Y and Z. Yunos, Ethics in Information Security, NST Tech & U, pp 1-4, November 2006

[9] Mohammad, S.S., Highlighting CAPTCHA,Human System Interactions, Conference on Digital Object Identifier, pp. 247-250. 2008.

[10] Applications of CAPTCHA, Key CAPTCHA, [online], Available from: https://www.keyCAPTCHA.com [20-Sep-2012].

[11] Sajad, S.S. and Ali, M. A New Anti-Spam Protocol Using CAPTCHA, proceedings of the IEEE International Conference on Networking, Sensing and Control (ICNSC07), London, United Kingdom, April 2007, pp. 234-238.

[12] Karrer, R. P. EC: an edge-based architecture against DDoS attacks and malware spread, Proceedings of 20th International Conference on Advanced Information Networking and Applications, vol.2, 2006, pp. 49-56.

[13] Sajad, S.S. Hossein, S. and Mohammad, S.S. Parental Control Based on Speaker Class Verification, IEEE Transactions on Consumer Electronics, vol. 54, no. 3, August 62008, pp. 1244-1251.

[14] Butler, W. L., Computer Security in the Real World, IEEE Journals&Magazines in Computer, Vol. 37, Issue 6. pp. 37-46. 2004.

[15] Yassin, Y and Z. Yunos, Ethics in Information Security, NST Tech & U, pp 1-4, November 2006.

[16] Masron, M.D.Z.I., Computer Security and Computer Ethics Awareness: A Component of Management Information System, International Symposium on Information Technology, pp. 1-7. 2008.

[17] Wen, Y.L., Young, A.L. and Ya, L.S., A Security Multi-Bank E-Cash Protocol Based on Smart Card, International Conference on Machine Learning and Cyberneties. pp. 3244-3248. 2007.

[18] HuyD, T., Christopher F. T. and CliffC, Z, iCAPTCHA: The Next Generation of CAPTCHA Designed to Defend Against 3rd Party Human Attacks, International conference on Communications (ICC)., pp 1-6, June 2011.

[19] Sajad, S.S. and Ali, M. A New Anti-Spam Protocol Using CAPTCHA, proceedings of the IEEE International Conference on Networking, Sensing and Control (ICNSC07), London, United Kingdom, April 2007, pp. 234-238.

[20] Sajad, S.S. and Ali, M., A New Anti-Spam Protocol Using CAPTCHA, Proceedings of the IEEE International Conference on Networking, Sensing and Control, pages 234-238, April 2007.

[21] Greg, M. and Jitendra, M. Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA, Proceedings of IEEE Conf. Computer Vision and Pattern Recognition, 2003. Voll. pp 134-141.

[22] Liu, S., Xing, Y. W., Xiao, M Y., Realization of 3D Animation Simulation and Scene Tour based on VC + + and OpenGL, Computer Engineering and Design., vol27, no.17, pp. 3235-3238, 2006.

[23] John, L.M. and Micheal, C., Visualization with OpenGL:3D Made Easy, IEEE Antennas and Propagation Magazine., vol 39, no. 4, pp. 116-120, August 1997.

[24] Abdulaziz S.A., Yasir, A., and Shouket A. K., Multi-Model CAPTCHA: A User Verification Scheme, International Conference on Information Science and Applications, pp. 1-7, 2011.

[25] Haichan G., et al. A Novel Image Based CAPTCHA Using Jigsaw Puzzle, 13th International Conference on Computational Science and Engineering., pp. 351-356, 2010.

[26] John, L.M. and Micheal, C., Visualization with OpenGL:3D Made Easy, IEEE Antennas and Propagation Magazine., vol 39, no. 4, pp 116-120, August 1997.

[27] Amalia, R. and Rebecca, D., Securing the Web Using Human Perception and Visual Object Interpretation, Proceeding of 13th International Conference Information Visualization, pp 613-618, July 2009.

[28] Chew, M. and Tygar, J.D., Image recognition CAPTCHAs Proceedings of the 7th International Information Security Conference (ISC 2004), pp 268–279, September 2004.

[29] Mohammad, S.S. and Sajad, S.S. Dynamic CAPTCHA, ICCSE internation Conference pp.676-679. 2012.

[30] M.H. Shirali-Shahreza and M. Shirali-Shahreza, Advanced Nastaliq CAPTCHA, 7th IEEE International Conference in Cybernetic Intelligent Systems, 2008.

[31] Mohammad, S.S., Highlighting CAPTCHA,Human System Interactions, Conference on Digital Object Identifier, pp. 247-250. 2008.

[32] Ahmad, E.A., Jeff, Y., and Wai, Y.N., CAPTCHA design: colour, usability and security, IEEE Internet Computing, pp. 1, July 2011.

[33] Sajad, S.S and Shirali, S., CAPTCHA for Blind People, IEEE International Symposium on Signal Processing and Information Technology, pp. 995-998, February 2008.

[34] Bhutan Demographics Profile 2013, Indexmundi, [online] Available from: http://www.indexmundi.com/bhutan/demographics_profile.html [March 2013].

[35] Top 10 Largest countries in the world, yahoo, [online] Available from: http://voices.yahoo.com/top-10-largest-countries-world-5807056.html [March 2013].

[36] Bhutan Internet Users, IndexMundi, [online] Available from: http://www.indexmundi.com/facts/bhutan/internet-users [March 2013].

[37] Literacy Rate in India, Maps of India, [online] Available from: http://www.mapsofindia.com/census2011/literacy-rate.html [March 2013].

[38] Thailand Population, Trading Economics, [online] Available from: http://www.tradingeconomics.com/thailand/population [March 2013].

[39] List of Countries by Literacy rate, Wikipedia Article, [online] Available from: http://en.wikipedia.org/wiki/List_of_countries_by_literacy_rate [March 2013].

[40] Internet Users in Thailand, Trading Economics, [online], Available from: http://www.tradingeconomics.com/thailand/internet-users-wb-data.html [March 2013].

APPENDIX

APPENDIX

EXPERIMENTAL RESULTS

Tables Below shows the Chi-square Test of correctness of the presented Text-based CAPTCHA in comparison with the nationalities of the users.

Table A1 Chi-square Test of Correctness of Character compared with Nationalities.

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	205.909[a]	2	.000
Likelihood Ratio	315.730	2	.000
N of Valid Cases	11406		

Below tables shows the correctness of the presented Text-based CAPTCHA in comparison with the nationalities of the users.

Table A2 Correctness of 1st Character compared with Nationalities.

		Correctness of 1st Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Nationality	Bhutanese	6550	510	7060	93%	7%
	Indian	2291	0	2291	100%	0%
	Thai	1989	66	2055	97%	3%
Total		10830	576	11406	95%	5%

Table A3 Correctness of 2nd Character compared with Nationalities.

		Correctness of 2nd Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Nationality	Bhutanese	6454	606	7060	91%	9%
	Indian	2291	0	2291	100%	0%
	Thai	1994	61	2055	97%	3%
Total		10739	667	11406	94%	6%

Table A4 Correctness of 3rd Character compared with Nationalities.

		Correctness of 3rd Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Nationality	Bhutanese	6455	605	7060	91%	9%
	Indian	2291	0	2291	100%	0%
	Thai	1978	77	2055	96%	4%
Total		10724	682	11406	94%	6%

Table A5 Correctness of 4th Character compared with Nationalities.

		Correctness of 4th Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Nationality	Bhutanese	6514	546	7060	92%	8%
	Indian	2291	0	2291	100%	0%
	Thai	1988	67	2055	97%	3%
Total		10793	613	11406	95%	5%

Table A63 Correctness of 5th Character compared with Nationalities.

		Correctness of 5th Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Nationality	Bhutanese	6486	574	7060	92%	8%
	Indian	2291	0	2291	100%	0%
	Thai	1964	91	2055	96%	4%
Total		10741	665	11406	94%	6%

Below tables shows the correctness of the presented Text-based CAPTCHA in comparison with the Age Group of the users.

Table A7 Correctness of 1st Character compared with Age Group.

		Correctness of 1st Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Age Group	10-25	2235	114	2349	95%	5%
	26-40	7729	400	8129	95%	5%
	41-60	809	61	870	93%	7%
	60 +	57	1	58	98%	2%
Total		10830	576	11406	95%	5%

Table A8 Correctness of 2nd Character compared with Age Group.

		Correctness of 2nd Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Age Group	10-25	2202	147	2349	94%	6%
	26-40	7676	453	8129	94%	6%
	41-60	804	66	870	92%	8%
	60 +	57	1	58	98%	2%
Total		10739	667	11406	94%	6%

Table A94 Correctness of 3rd Character compared with Age Group.

		Correctness of 3rd Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Age Group	10-25	2194	155	2349	93%	7%
	26-40	7669	460	8129	94%	6%
	41-60	804	66	870	92%	8%
	60 +	57	1	58	98%	2%
Total		10724	682	11406	94%	6%

Table A105 Correctness of 4th Character compared with Age Group.

		Correctness of 4th Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Age Group	10-25	2230	119	2349	95%	5%
	26-40	7694	435	8129	95%	5%
	41-60	811	59	870	93%	7%
	60 +	58	0	58	100%	0%
Total		10793	613	11406	95%	5%

Table A11 Correctness of 5th Character compared with Age Group.

		Correctness of 5th Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Age Group	10-25	2205	144	2349	94%	6%
	26-40	7681	448	8129	94%	6%
	41-60	798	72	870	92%	8%
	60 +	57	1	58	98%	2%
Total		10741	665	11406	94%	6%

Below tables shows the correctness of the presented Text-based CAPTCHA in comparison with the Career or Educational Back Group of the users.

Table A12 Correctness of 1st Character compared Career/Education Back Ground.

		Correctness of 1st Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Career	Corporate/Private Organization	5605	245	5850	96%	4%
	Government/State Official	2271	243	2514	90%	10%
	Trade/Business	885	13	898	99%	1%
	Others	694	31	725	96%	4%
	Student	1296	36	1332	97%	3%
	House Wife/Retiree	79	8	87	91%	9%
Total		10830	576	11406	95%	5%

Table A13 Correctness of 2nd Character compared Career/Education Back Ground.

		Correctness of 2nd Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Career	Corporate/Private Organization	5544	306	5850	95%	5%
	Government/State Official	2250	264	2514	89%	11%
	Trade/Business	883	15	898	98%	2%
	Others	690	35	725	95%	5%
	Student	1289	43	1332	97%	3%
	House Wife/Retiree	83	4	87	95%	5%
Total		10739	667	11406	94%	6%

Table A14 Correctness of 3rd Character compared Career/Education Back Ground.

		Correctness of 3rd Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Career	Corporate/Private Organization	5535	315	5850	95%	5%
	Government/State Official	2264	250	2514	90%	10%
	Trade/Business	881	17	898	98%	2%
	Others	691	34	725	95%	5%
	Student	1273	59	1332	96%	4%
	House Wife/Retiree	80	7	87	92%	8%
Total		10724	682	11406	94%	6%

Table A15 Correctness of 4th Character compared Career/Education Back Ground.

		Correctness of 4th Character			Percentage	
		Correct	Not Correct	Total	Correct	Not Correct
Career	Corporate/Private Organization	5555	295	5850	95%	5%
	Government/State Official	2288	226	2514	91%	9%
	Trade/Business	889	9	898	99%	1%
	Others	689	36	725	95%	5%
	Student	1288	44	1332	97%	3%
	House Wife/Retiree	84	3	87	97%	3%
Total		10793	613	11406	95%	5%

Table A16 Correctness of 5th Character compared Career/Education Back Ground.

		Correctness of 5th Character		Total	Percentage	
		Correct	Not Correct		Correct	Not Correct
Career	Corporate/Private Organization	5556	294	5850	95%	5%
	Government/State Official	2266	248	2514	90%	10%
	Trade/Business	883	15	898	98%	2%
	Others	680	45	725	94%	6%
	Student	1273	59	1332	96%	4%
	House Wife/Retiree	83	4	87	95%	5%
Total		10741	665	11406	94%	6%

Below tables shows the correctness of the presented Text-based CAPTCHA in comparison with the Working in Computer of the users.

Table A17 Correctness of 1st Character compared Working in Computer.

		Correctness of 1st Character		Total	Percentage of Correctness of 1st Character	
		Correct	Not Correct		Correct	Not Correct
Working in Computers	No	2311	183	2494	93%	7%
	Yes	8519	393	8912	96%	4%
Total		10830	576	11406	95%	5%

Table A18 Correctness of 2nd Character compared Working in Computer.

		Correctness of 2nd Character			Percentage of Correctness of 2nd Character	
		Correct	Not Correct	Total	Correct	Not Correct
Working in Computers	No	2255	239	2494	90%	10%
	Yes	8484	428	8912	95%	5%
Total		10739	667	11406	94%	6%

Table A196 Correctness of 3rd Character compared Working in Computer.

		Correctness of 3rd Character			Percentage of Correctness of 3rd Character	
		Correct	Not Correct	Total	Correct	Not Correct
Working in Computers	No	2250	244	2494	90%	10%
	Yes	8474	438	8912	95%	5%
Total		10724	682	11406	94%	6%

Table A20 Correctness of 4th Character compared Working in Computer.

		Correctness of 4th Character			Percentage of Correctness of 4th Character	
		Correct	Not Correct	Total	Correct	Not Correct
Working in Computers	No	2294	200	2494	92%	8%
	Yes	8499	413	8912	95%	5%
Total		10793	613	11406	95%	5%

Table A21 Correctness of 5th Character compared Working in Computer.

		Correctness of 5th Character			Percentage of Correctness of 5th Character	
		Correct	Not Correct	Total	Correct	Not Correct
Working in Computers	No	2296	198	2494	92%	8%
	Yes	8445	467	8912	95%	5%
Total		10741	665	11406	94%	6%

Table A22 Correctness of 1st Character compared visibility of user

		Correctness of 1st Character			Correctness of 1st Character	
		Correct	Not Correct	Total	Correct	Not Correct
Visibility	Normal	8500	443	8943	95%	5%
	Short	1132	84	1216	93%	7%
	Long	917	40	957	96%	4%
	Short and Long Tilt	111	5	116	96%	4%
	Short Inclined	170	4	174	98%	2%
Total		10830	576	11406	95%	5%

Table A23 Correctness of 2nd Character compared visibility of user

		Correctness of 2nd Character			Percentage of Correctness of 2nd Character	
		Correct	Not Correct	Total	Correct	Not Correct
Visibility	Normal	8410	533	8943	94%	6%
	Short	1129	87	1216	93%	7%
	Long	916	41	957	96%	4%
	Short & Long Tilt	114	2	116	98%	2%
	Short Inclined	170	4	174	98%	2%
Total		10739	667	11406	94%	6%

Table A24 Correctness of 3rd Character compared visibility of user

		Correctness of 3rd Character			Percentage of Correctness of 3rd Character	
		Correct	Not Correct	Total	Correct	Not Correct
Visibility	Normal	8390	553	8943	94%	6%
	Short	1136	80	1216	93%	7%
	Long	914	43	957	96%	4%
	Short & Long Tilt	114	2	116	98%	2%
	Short Inclined	170	4	174	98%	2%
Total		10724	682	11406	94%	6%

Table A25 Correctness of 4th Character compared visibility of user

		Correctness of 4th Character			Percentage of Correctness of 4th Character	
		Correct	**Not Correct**	**Total**	**Correct**	**Not Correct**
Visibility	Normal	8457	486	8943	95%	5%
	Short	1145	71	1216	94%	6%
	Long	914	43	957	96%	4%
	Short and Long Tilt	107	9	116	92%	8%
	Short Inclined	170	4	174	98%	2%
Total		10793	613	11406	95%	5%

Table A26 Correctness of 5th Character compared visibility of user

		Correctness of 5tht Character			Percentage of Correctness of 5th Character	
		Correct	**Not Correct**	**Total**	**Correct**	**Not Correct**
Visibility	Normal	8405	538	8943	94%	6%
	Short	1145	71	1216	94%	6%
	Long	918	39	957	96%	4%
	Short & Long Tilt	106	10	116	91%	9%
	Short Inclined	167	7	174	96%	4%
Total		10741	665	11406	94%	6%

Tables below show the correctness of Text-based CAPTCHA with Career of users along with the age-group of the users.

Table A27 Correctness of 1st Character with Career and age-group of the users.

Age Group			Correctness of 1st Character		Total	Correctness of 1st Character		Total
			Correct	Not Correct		Correct	Not Correct	
10-25	Career	Govt./State official	345	32	377	92%	8%	100%
26-40	Career	Govt./State official	1751	183	1934	91%	9%	100%
41-60	Career	Govt./State official	147	27	174	84%	16%	100%
60 +	Career	Govt./State official	28	1	29	97%	3%	100%
	Total		2271	243	2514	90%	10%	100%

Table A28 Correctness of 2nd Character with Career and age group of users.

Age Group			Correctness of 2nd Character		Total	Percentage of Correctness of 2nd Character		Total
			Correct	Not Correct		Correct	Not Correct	
10-25	Career	Govt./State official	342	35	377	91%	9%	100%
26-40	Career	Govt./State official	1736	198	1934	90%	10%	100%
41-60	Career	Govt./State official	144	30	174	83%	17%	100%
60 +	Career	Govt./State official	28	1	29	97%	3%	100%
	Total		2250	264	2514	97%	3%	100%

Table A29 Correctness of 3rd Character with Career and age-group of user.

Age Group			Correctness of 3rd Character		Total	Percentage of Correctness of 3rd Character		Total
			Correct	Not Correct		Correct	Not Correct	
10-25	Career	Govt./State	344	33	377	91%	9%	100%
26-40	Career	Govt./State	1750	184	1934	90%	10%	100%
41-60	Career	Govt./State	142	32	174	82%	18%	100%
60 +	Career	Govt./State	28	1	29	97%	3%	100%
	Total		2264	250	2514	97%	3%	100%

Table A30 Correctness of 4th Character with Career and age-group of users.

Age Group			Correctness of 4th Character		Total	Percentage of Correctness of 4th Character		Total
			Correct	Not Correct		Correct	Not Correct	
10-25	Career	Govt./State	357	20	377	95%	5%	100%
26-40	Career	Govt./State	1756	178	1934	91%	9%	100%
41-60	Career	Govt./State	146	28	174	84%	16%	100%
60 +	Career	Govt./State	29	0	29	100%	0%	100%
	Total		2288	226	2514	100%	0%	100%

Table A31 Correctness of 5th Character with Career and age group of users.

Age Group			Correctness of 5tht Character		Total	Percentage of Correctness of 5th Character		Total
			Correct	Not Correct		Correct	Not Correct	
10-25	Career	Govt./State	344	33	377	91%	9%	100%
26-40	Career	Govt./State	1758	176	1934	91%	9%	100%
41-60	Career	Govt./State	136	38	174	78%	22%	100%
60 +	Career	Govt./State	28	1	29	97%	3%	100%
	Total		2266	248	2514	97%	3%	100%

Tables below show the correctness of Text-based CAPTCHA with Career of users along with the working habits of the users.

Table A32 Correctness of 1st character with Career and Working habits.

Working in Computers			Correctness of 1st Character		Total	Percentage of Correctness of 1st Character		Total
			Correct	Not Correct		Correct	Not Correct	
No	Career	Govt./State	630	95	725	87%	13%	100%
Work in Computer	Career	Govt./State	1641	148	1789	92%	8%	100%
	Total		2271	243	2514	90%	10%	100%

Table A33 Correctness of 2nd character with Career and Working habits.

Working in Computers			Correctness of 2nd Character			Percentage of Correctness of 2nd Character		
			Correct	Not Correct	Total	Correct	Not Correct	Total
No	Career	Govt./State	613	112	725	85%	15%	100%
Work in Computer	Career	Govt./State	1637	152	1789	92%	8%	100%
	Total		2250	264	2514	89%	11%	100%

Table A34 Correctness of 3rd character with Career and Working habits.

Working in Computers			Correctness of 3rd Character			Percentage of Correctness of 3rd Character		
			Correct	Not Correct	Total	Correct	Not Correct	Total
No	Career	Govt./State	624	101	725	86%	14%	100%
Work in Computer	Career	Govt./State	1640	149	1789	92%	8%	100%
	Total		2264	250	2514	90%	10%	100%

Table A35 Correctness of 4th character with Career and Working habits.

Working in Computers			Correctness of 4th Character		Total	Percentage of Correctness of 4th Character		Total
			Correct	Not Correct		Correct	Not Correct	
No	Career	Govt./State	635	90	725	88%	12%	100%
Work in Computer	Career	Govt./State	1653	136	1789	92%	8%	100%
	Total		2288	226	2514	91%	9%	100%

Table A36 Correctness of 5th character with Career and Working habits.

Working in Computers			Correctness of 5tht Character		Total	Percentage of Correctness of 5tht Character		Total
			Correct	Not Correct		Correct	Not Correct	
No	Career	Govt./State	631	94	725	87%	13%	100%
Work in Computer	Career	Govt./State	1635	154	1789	91%	9%	100%
	Total		2266	248	2514	90%	10%	100%

BIOGRAPHY

Tsheten Tamang was born on 5th August 1984 in Chargharey, Samtse, Bhutan. He graduated in Bachelor in Computer Application (BCA) from Patrician College of Art and Science, Affiliated to University of Madras, Chennai – India, in the year 2006. He worked for Drukonnet Pvt. Ltd. Thimphu, Bhutan, for 2 years (2006 - 2008) as Network Administrator. 2008 – Till date he is working for National Pension and Provident Fund, Thimphu, Bhutan, as Program Officer, ICT.

Master in Computer Science and Information Technology, from Chulalongkorn University, Bangkok, Thailand, in 2013, under the scholarship between Thailand International Development Cooperation Agency (TICA) of Thailand and National Pension and Provident Fund, Thimphu, Bhutan.

Publication Paper : Tsheten T., and Bhattarakosol P. Uncover Impact Factors of Text-based CAPTCHA Identification. Computing and Convergence Technology (ICCIT, ICEI and ICACT), Proceeding of 2012 7th International Conference 2012, pp. 556-560, South Korea: IEEE, 2012.