REPUBLIC OF TURKEY
FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

AN ANALYSIS OF THE STRENGTH OF CAPTCHA USED BY SOME TURKISH UNIVERSITIES TO PROTECT STUDENT INFORMATION

KARMAND HUSSEIN ABDALLA

Master Thesis
Department: Software Engineering
Supervisor: **Asst. Prof. Dr. Mehmet KAYA**

JULY-2017

REPUBLIC OF TURKEY
FIRAT UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

AN ANALYSIS OF THE STRENGTH OF CAPTCHA USED BY SOME TURKISH UNIVERSITIES TO PROTECT STUDENT INFORMATION

MASTER THESIS

KARMAND HUSSEIN ABDALLA
(Student No: 151137114)

Department of Software Engineering

Submission Date to the Institute: June 7, 2017
Thesis Presentation Date: July 4, 2017

Thesis Supervisor: Asst. Prof. Dr. Mehmet KAYA (ADYU)

Jury member: Prof. Dr. Asaf VAROL (FU)

Jury member: Assoc. Prof. Dr. Cihan VAROL (SHSU, USA)

July-2017

ACKNOWLEDGMENT

I am much obliged to my supervisor, Assistant Professor Dr Mehmet Kaya, who provided such a great amount of help in my time at university. I wanted to let him know that I have arrived at Firat University studying Software Engineering, and that I couldn't have settled on this choice without his assistance. I would also like to thank him for his kind help and advice while I was working in this thesis, without his assistance was excessively troublesome, making it impossible to complete it.

I also wish to give many thanks to my supervisor for his tutelage, exhortation, and direction amid my since study at Firat University. Working with him has as of now helped me so much and will keep on inspiring me. His commitment and dedication with his time in helping me with other side-undertakings are enormously refreshing. It is my delight to send him my gratefulness.

He has unquestionably been an incredible tutor for me and he has been cooperative in sharing his rich and commendable learning in your field!

I would like to thank all my friends, who have supported me throughout the entire process, both by keeping me harmonious and helping me putting pieces together. I will be grateful forever for, besides all those peoples which I thanked them in the beginning I want to show very thankfully to my wife which she was always beside me while doing any projects in my life.

Last but not least, I want to thank my family: my mother, father, brothers and sister for supporting me profoundly all through composing this proposal and my life in general.

Karmand Hussein

Elazig – July/2017

TABLE OF CONTENTS

Page No

ACKNOWLEDGMENT II

TABLE OF CONTENTS III

ABSTRACT V

ÖZET VI

LIST OF FIGURE VII

ABBREVIATIONS IX

1. INTRODUCTION 1

1.1. Overview 1

1.2. Literature Review 3

1.3. Problem Statement 6

1.4. Aim of the Thesis 6

1.5. Thesis Layout 7

2. BACKGROUND AND DETAILED EXPLANATION OF CAPTCHA TYPES 8

2.1. Introduction 8

2.1.1. Security Attacks 8

2.1.1.1. Active Attack 9

2.1.1.2. Passive Attack 10

2.1.2. Security Properties 11

2.1.3. Confidentiality 11

2.1.3.1. Integrity 12

2.1.3.2. Availability 12

2.1.4. Security Mechanisms 12

2.1.4.1. Firewall 13

2.1.4.2. VPN 13

2.1.4.3. UTM 14

2.1.4.4. IDS/IPS 14

2.2. Detailed Explanation of CAPTCHA Types 15

2.2.1. Text-based CAPTCHA 15

2.2.2. Image-based CAPTCHA 18

2.2.3. Video-based CAPTCHA 22

2.2.4. Audio-based CAPTCHA 23

2.2.5. Puzzle-based CAPTCHA 24
2.2.6. Question-based CAPTCHA 25
2.3. CAPTCHA Drawbacks 28
2.4. A CAPTCHA Recognition Approach 28
2.4.1. Pre-processing 29
2.4.2. Number Splitting Procedure 29
2.4.3. The Probability Pattern Module 30
2.4.4. Number Recognition CAPTCHA Image 31
2.5. Breaking CAPTCHAs 31
3. PROPOSED TECHNIQUES 34
3.1. Website Prototyping 34
3.1.1. Question-based CAPTCHA Generation 34
3.1.2. BCAPTCHA Database 35
3.2. Browser Add-on 35
3.3. The CAPTCHA Recognition by Proposed OCR System 36
3.3.1. Noise Reduction 37
3.3.2. Grayscale 38
3.3.3. Binarisation 39
3.3.4. Median Filter 40
3.4. Implementing OCR Process 41
3.5. Dictionary Attack 42
4. EVALUATION AND RESULTS 43
4.1. Evaluation 43
4.2. Results 45
5. CONCLUSIONS AND RECOMMENDATIONS 47
6. REFERENCES 48
CURRICULUM VITA 54

ABSTRACT

AN ANALYSIS OF THE STRENGTH OF CAPTCHA USED BY SOME TURKISH UNIVERSITIES TO PROTECT STUDENT INFORMATION

With increasing Internet usage worldwide, the need to protect websites from attackers is paramount. As so many people perform tasks such as shopping and banking, developers must pay close attention to the security of their sites. A significant vulnerability can exist during the client service registration process where users register for a free Internet service such as an email account. During enrollment, a few interlopers or aggressors may carry out malicious attacks by enlisting bots to create fake accounts and drain or waste site resources. Security analysts have created various systems to protect web assets from these bots. To differentiate a human user from a bot, CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is a commonly employed strategy. This study focuses on the use of such a strategy in the field of web security to mitigate attacks. Different text, image, audio and video-based CAPTCHAs are investigated including those that are based on puzzles.

This thesis sheds light on the type of CAPTCHA used by some Turkish universities to protect students' specific information pages. It is a mixture of two types of CAPTCHA: text and image. After examining this type of CAPTCHA, it is broken and recognised. This is achieved taking into consideration some weak points in its structure, which are discussed in the results section. The process of breaking and recognising is done via an add-on and a program in Java language. Then, by using OCR technique, the weakness of this CAPTCHA can be proven. In 53 CAPTCHA cases, 92.45% can be broken and recognised.

Keywords: CAPTCHA breaking, OCR, security, strength, character recognition.

ÖZET

TÜRK ÜNİVERSİTELERİNİN ÖĞRENCİ BİLGİLERİNİ KORUMAK İÇİN KULLANDIKLARI CAPTCHA SİSTEMİNİN BİR ANALİZİ

İnternet kullanımının dünya çapında artmasıyla, web sitelerini saldırganlardan korumanın gereği de daha çok önem kazanmıştır. Birçok kişi artık alışveriş ve bankacılık gibi işlemleri internet üzerinden gerçekleştirdiğinden, geliştiriciler sitelerinin güvenliğine yakından dikkat etmelidir. Kullanıcıların e-posta hesabı gibi ücretsiz internet servislerine kaydolduğu istemci hizmet kaydı işlemi sırasında önemli güvenlik açıkları oluşabilir. Bu tip kayıtlar sırasında, saldırganlar bot denen programlar oluşturup sahte hesaplar oluşturarak web sitesinin kaynaklarını kötü niyetli saldırılara maruz bırakabilirler. Güvenlik analistleri, web varlıklarını bu botlardan korumak için çeşitli sistemler oluşturmuştur. Bir insandan bir botu ayırt etmek için, CAPTCHA (İnsan ve Bilgisayar Ayrımı Amaçlı Tam Otomatik Genel Turing Testi) sık kullanılan bir stratejidir. Bu çalışma, saldırıları azaltmak için web güvenliği alanında bu stratejinin kullanılmasına odaklanmaktadır. Bulmacalara dayanan metinler de dahil olmak üzere, farklı metin, resim, ses ve video tabanlı CAPTCHA'lar araştırılmıştır.

Bu tez, bazı Türk üniversitelerinin öğrencilerin özel bilgi sayfalarını korumak için kullandıkları CAPTCHA türüne ışık tutmaktadır. CAPTCHA nin metin ve resimedn olusan iki tipinin karişimidir. Bu CAPTCHA tipi inceledikten sonra kırılmış ve tanınmıştır. Bu, sonuçlar bölümünde tartıştığımız, bu tür CAPTCHAlerin yapısındaki bazı zayıf noktalar göz önüne alınarak başarılmıştır. Kırma ve tanıma süreci için bir Mozilla eklentisi kullanılmış ve Java programı oluşturulmuştur. Ardından, OCR tekniği kullanılarak, bu CAPTCHA'nın zayıflığı kanıtlanmıştır. Kullandığımız 53 CAPTCHA örneğinden% 92.45'inin kırılabilir ve tanınabilir olduğu gösterilmiştir.

Anahtar Kelimeler: CAPTCHA kırma, OCR, güvenlik, güç, karakter tanıma.

LIST OF FIGURE

Page No

Figure 1.1. CAPTCHA test simple. 3
Figure 2.1. CAPTCHA. 15
Figure 2.2. Text-based CAPTCHA 16
Figure 2.3. Simple CAPTCHAs 17
Figure 2.4. Math CAPTCHA. 17
Figure 2.5. Text Based on CAPTCHA 18
Figure 2.6. Image-based CAPTCHA 19
Figure 2.7. Image-based CAPTCHA. 20
Figure 2.8. Layered text and colour combination CAPTCHA 20
Figure 2.9. Option image CAPTCHA 21
Figure 2.10. Rotate image CAPTCHA 21
Figure 2.11. Ordering image CAPTCHA 22
Figure 2.12. Video-based CAPTCHA 23
Figure 2.13. Audio-based CAPTCHA 24
Figure 2.14. Puzzle-based CAPTCHA. 25
Figure 2.15. Choose the correct answer 26
Figure 2.16. Question-based CAPTCHA. 26
Figure 2.17. Colour image-based CAPTCHA. 27
Figure 2.18. Multi-colour image-based CAPTCHA 27
Figure 2.19. Bongo CAPTCHA [16]. 28
Figure 2.20. a-b (original images), c-d (grey-level images), e-f (binary images), g-h (after noise reduction) 29
Figure 2.21. Split the numbers 29
Figure 2.22. a (-45 to 0 degree tilted) and b (0 to 45 degree tilted) 30
Figure 2.23. The sample of suggested CAPTCHA recognition 31
Figure 2.24. Characters with same or similar projection features. 32
Figure 2.25. Guide lines principle 33
Figure 2.26. Detect G and L (Left), Detect H (Right). 33
Figure 3.1. The prototyped website (http://www.bcaptcha.com). 34

Figure 3.2. Pre-processing noise reduction flow chart diagram. 37
Figure 3.3. The image before processing..... 38
Figure 3.4. The image after noise reduction. 38
Figure 3.5. Applying thresholding technique to a contrasted and grayscale image. 39
Figure 3.6. Effects of applying median filter technique to a noisy image. 40
Figure 3.7. Image noise reduction and its OCR-produced equation..... 41
Figure 3.8. Run-time process. 41
Figure 3.9. Login page during software attack: entering username and password for the correct CAPTCHA..... 42
Figure 4.1. The diagram shows a system of recognition framework for question-based CAPTCHAs 43
Figure 4.2. Clicking on the check box pass of the reCAPTCHA. 44
Figure 4.3. Google new reCAPTCHA. 45

ABBREVIATIONS

AMF : Adaptive Median Filter

CBPA : CAPTCHA Based Password Authentication

CHT : Cut Head Tail

CIA : Confidentiality Integrity Availability

DDoS : Distributed Denial of Service

IDS : Intrusion Detection System

NLP : Natural Language Processing

OCR : Optical Character Recognition

UTM : Unified Threat Management

VPN : Virtual Private Network

1. INTRODUCTION

1.1. Overview

The term CAPTCHA stands for Completely Automated Public Turing Test to Tell Computers and Humans Apart. The real test of an effective CAPTCHA test is to build a program which can make and grade challenges, in order to permit a great number of people, but prohibit PCs (bots). This study proposes another CAPTCHA component which depends on customisation of different security parameters, for example, contortion, straightforwardness, character set and length, to build a more secure alternative to the current frameworks.

While utilising a Gmail account, using PayPal or possibly leaving a comment on some notable web journals, a user is likely to come across a CAPTCHA in some shape or form [1]. Put simply, a CAPTCHA is a security system that is utilised to forestall independent sections on sites (or anticipate mechanised bots) from performing unlawful exercises on pages.

Gmail enhances its administration by blocking access to computerised spammers that perform auto recruits of Google accounts, while eBay delays bots by inundating the location with puzzles. Similarly, Facebook limits the production of fake profiles used to spam and stalk legitimate clients or cheat [2]. These are all genuine cases of CAPTCHAs.

While using a Gmail account or PayPal, CAPTCHA utilises a mix of distorted characters and obscure systems that people can perceive. However, this is sufficiently troublesome for computerised scripts. CAPTCHA, also known as the "Turn around Turing Tests," enables a PC to decide whether the distant customer is human [2]. However, bots have developed into more up-to-date instruments to discover escape clauses in the current CAPTCHA frameworks. Thus, CAPTCHAs must be redesigned from time-to-time to resist these assaults.

The most widely recognised use of CAPTCHAs today is to ensure websites' safety from automated registrations or other spam activities. Numerous businesses (Microsoft, Yahoo, etc.) have free email facilities. Until a couple of years ago, the majority of these organisations experienced a particular kind of attack, whereby bots and automated programs

would agree to accept a huge number of email records [3]. The answer was to use CAPTCHAs to guarantee that only people could acquire free accounts, in order to protect website resources. As a general rule, free administration ought to be ensured with a CAPTCHA, keeping in mind the goal to counteract bots [4]. Furthermore, CAPTCHAs are useful for several other similar applications, such as preventing website registration, protection from spam comments in blogs and stopping dictionary attacks in password systems [5].

The literature shows that for successful implementation and effective use, a CAPTCHA must satisfy some fundamental requirements. For example, it must be simple enough for a human user to pass, and at the same time, sufficiently difficult for a bot [6]. On the other hand, one must consider the user friendliness of the website while designing and embedding a CAPTCHA. In this section, the different types of CAPTCHAs are compared and their effectiveness evaluated.

A CAPTCHA is used with graphical keys for speech spyware. A content CAPTCHA is shown underneath the images, and the user finds his individual pass pictures and enters the characters in particular areas as a key for confirmation. Interactive focuses in Text Points are notable for their character, and consequently, they recall a password. However, they cannot be misused by machines (when compared with still focuses in conventional graphical watchword plans) and logical, active areas of interactive focuses on their connections. Text points pictures appear to be indistinguishable from Click Text pictures and are produced similarly. The areas of all the interactive focuses are checked to guarantee that none is blocked or that its resistance district covers another interactive point. CAPTCHA in authentication can be used with a password in client confirmation convention. The CAPTCHA is based on Password Authentication (CbPA) convention to counter online lexicon assaults. Pass Points is for the most part considered a snap based survey arrangement where a user snaps progression of spotlights where a photo is making a watchword, and re-taps a similar course of action during confirmation [7].

Although most people are not familiar with the term CAPTCHA, many use it on a weekly or even daily basis. CAPTCHAs are the security tests that are most often found on websites that require the use of registration [8]. These security tests often consist of a series of distorted characters that the user must type in order to continue to the desired location. The purpose of the CAPTCHA is to ensure that access to a website or other resource is being

performed by a human rather than an automated computer program (automated attack). By using CAPTCHA tests, all attackers are resisted. For instance, people can read misshaped content; however, PC programs (bot and automated) may not be able to pass. The term "Turing Test" refers to a game in which a human judge may ask a number of questions without knowing which player is the computer and which is the human [9,10]. The judge, who must make this distinction, is the computer player. The judge is no longer a human, but is instead the computer server hosting the CAPTCHA; this computer must determine whether the user is human or a computer [11].

A CAPTCHA is a security instrument requiring the correct response to a question, which only a human can answer as opposed to any arbitrary figure. People have speed constraints, and consequently, cannot repeat the effect of a robotised program [4]. Hence, the fundamental prerequisite of a CAPTCHA is that PC programs must be slower than people in reacting accurately, because the responses of the automated programs are wrong or they face difficulties in providing answers. Additionally, a CAPTCHA is useful for several applications, such as preventing website registration and leaving spam comments in blogs, stopping email addresses from web scammers' online polls, and avoiding worms and spam in emails and dictionary attacks in a password system [12].

Figure 1.1. CAPTCHA test simple.

1.2. Literature Review

This section considers the evolution of CAPTCHA. It was first used in 1997 to block automated URL submissions [13]. AltaVista Chief Scientist Andrei Broder and his partners

built a network which only individuals could read. As a result, spam and automated programs were reduced by 95% and a patent was issued in 2001 [14].

In November 1999, Slashdot.com reviewed the top CS colleges in the US. The Massachusetts Institute of Technology and Understudies from the Carnegie Mellon University completed machines to vote for their respective colleges. This made it necessary to use CAPTCHAs for these kinds of polls in order to avoid bots [15].

In 2000, the prominent Messenger service of YAHOO was hit by bots, which created much inconvenience for users. Hurray, alongside Carnegie Mellon University, developed EZ-GIMPY CAPTCHA; this would select a name and corrupt it using a wide range of images, and request from the user to incorporate the reshaped text [9].

In the last decade, there have been many attempts to prevent attackers, who continually seek new techniques to penetrate network services. Moreover, the web-developers try to find new forms of protection against attackers. Many different approaches have been proposed to make tests more effective, such as text, audio, image and video-based CAPTCHAs. The attackers of course also employed new methods to break the CAPTCHA tests. Some of the most recent CAPTCHA tests are image orientated, puzzles and using distorted text or mixing images with texts [12].

Text-based CAPTCHAs are weak because they can be broken [16]. This method utilises certain procedure, for example, pre-processing, division, post-division, acknowledgment and post-preparing, and then perceives the characters depending upon its elements [16].

Rahman [12] reviews the literature on CAPTCHA tests. He discusses the typical applications of CAPTCHA and its weaknesses and strengths based on text and image. The advantage of using image-based CAPTCHA is that the sample identification of a picture is unbreakable (AI problem). It is difficult to crack this test utilising design acknowledgment strategy. The shortcomings of the image-based CAPTCHA test is that it is accessible only in English. However, just 27% of users are English speakers. It is also problematic for users who have sight impairment or learning difficulties [12].

According to the article proposed by Choudharyit [17], it is not easy for PCs to separate the content. However, CAPTCHAs are for the most part basic tests and humans can identify the corruptions. Therefore, the right answers are needed to allow human users into the site. Moreover, clarifying content is based on CAPTCHAs. The least difficult yet novel approach

is to give a user a test that only humans can understand; for example, (i) What is 15 - 4?, (ii) Select the fourth letter in UNIVERSITY, (iii) Which of the words red, Friday and Ali is a colour?, (iv) If today is Saturday, when is Sunday?

Fisk [18] described a new type of CAPTCHA in his article, which arranges pictures in an artificial way. For example, when users wish to enter a website, the CAPTCHA first shows them three pictures, which they must arrange in order. For instance, these could be pictures of an egg, a hen and a rooster. The user knows that the egg should come first, then the hen and, finally, the rooster. The user can, subsequently, enter the website. However, an automated program cannot understand this process, and it will be prevented from entering the website.

A paper proposed by Singh and Pal [7], in their study of different kinds of CAPTCHAs, discuss the requests and problems associated with dissimilar types of tests. For example, in the text-based CAPTCHA test, a legitimate user may encounter some problems with identifying the right content or typescripts due to font size, multiple fonts, Wave Motion and Blurred Letter. However, these types of texts can be easily identified by a well-trained OCR algorithm. Similar problems may emerge with CAPTCHAs based on images for different reasons; for example, legitimate users may have low vision or see blurred images.

Another CAPTCHA technique is based on audio. Yet, this is not very common because most of the time it is available only in English. Therefore, this implies that to access websites protected by audio-based CAPTCHAs, users must be fluent in English in order not to experience problems with audio hardware. Furthermore, CAPTCHAs based on video are not practical, and they are difficult to implement due to the large size of video files. With video-based CAPTCHAs, users may also face other problems, for example, to download the video and solve the CAPTCHA.

Samruddhi [19] in his article introduced different types of CAPTCHA system in view of simplified mouse movement (DnD Captcha). They survey CAPTCHAs and compare the different techniques. The drag and drop CAPTCHA is proposed in response to the advanced optical character recognition (OCR) techniques for reading text-based CAPTCHAs. Their approach relies on the DnD CAPTCHA strategy to tell humans and PCs apart. In this type of CAPTCHA, the task is simply to drag and drop character blocks into their various blank blocks instead of typing a text.

1.3. Problem Statement

In this study, different CAPTCHAs are analysed. The focus, however, is on a commonly used type: the question-based CAPTCHA. Its strengths and weaknesses are examined. In addition, it is evaluated how well CAPTCHAs protect websites. In the case study, the focus is on question-based CAPTCHAs employed in the websites of some Turkish universities such as Adiyaman University, Harran University, Bingol University and Siirt University. Different methods and algorithms are investigated to understand how the websites mentioned can be strengthened.

1.4. Aim of the Thesis

The aim of this research is to gain insight into how each type of CAPTCHA is designed to secure websites. The study is concerned with breaking a picture-based CAPTCHA, specifically, a question-based CAPTCHA. The goal is to demonstrate that the images of question-based CAPTCHAs are not as secure as their creators assert. This, consequently, highlights the weaknesses of the distinctive applications utilising picture CAPTCHAs, especially question-based CAPTCHAs. The discussion underlines concerns over the security strategies created to protect privacy in online frameworks. Question-based CAPTCHAs, which represent the most recent advancements, are defenceless and susceptible to assaults by programmers with high achievement rates.

1. In order to achieve the main objective, the images in CAPTCHA tests should be contorted and altered so that they cannot be interpreted by OCR techniques easily. The first thing that can be done is to include noise in the picture, especially in the background, to make the automated extraction of text difficult. Therefore, the first step in CAPTCHA recognition is pre-processing to eliminate such noise. After converting the image to binary via Otsu's method, due to its relatively higher performance compared with others, the versatile median filter is utilised to kill commotion.

2. After the pre-processing, the next step is to split the numbers. For this, a projection method is employed. The black pixels are projected on the x-axis and the coordinates of each digit (0-9) are determined in the CAPTCHA image. It should be noted that the projection of multiple digits may overlap in some cases.

1.5. Thesis Layout

The thesis consists of four further chapters:

Chapter Two: [Background and detailed explanation of CAPTCHA types]

In this chapter, the different types of security and classifications of CAPTCHAs are evaluated based on their popularity and strength.

Chapter Three: [Proposed techniques]

In this chapter, the structure of the proposed technique for breaking the question-based CAPTCHA that has been used in some Turkish universities is analysed, and the steps taken to bypass the test by using OCR are described. Furthermore, the stages for implementing the proposed approach for recognising the extracted question from the image and estimating the results are described.

Chapter Four: [Evaluation and results]

This chapter examines the different types of CAPTCHAs, especially those used by some Turkish universities, the results of the research and the 53 different types of question-based CAPTCHAs.

Chapter Five: [Conclusions and recommendations]

The results of this work are summarised and future research recommended.

2. BACKGROUND AND DETAILED EXPLANATION OF CAPTCHA TYPES

2.1. Introduction

There has been a significant growth in Internet usage, and with improvement in the vast open systems, security dangers have increased in the last 20 years. Hence, there is constant focus on network security and on protecting all the services, including social media, bank systems and websites [9, 12]. Internet services help improve the daily lives of people in different places. They also manage sales and other forms of transactions. For this reason, protecting Internet services against attackers has become a necessity. It is vital to focus on the administration utilised through the web as there are many malignant forces on the Internet which may break a framework without any safe application. Security gives assurance against such dangers [20]. One such danger is the bot, which is a malevolent program or computerised program. It has the capacity to run robotised errands over the system and can create issues for a network. There are layers in the network and each has mechanisms for protecting against attacks. The most significant layer in the network is the Application Layer Network. Many attacks against a network use this layer because it can be done easily by manual and automated system. Protecting the network in the Application Layer Network prevents threats and, in so doing, stabilises the network to provide services [21].

2.1.1. Security Attacks

Previously, programmers comprehended the points of interest of PC interchanges and how to compromise vulnerabilities. Nowadays, practically anybody can become a programmer by downloading tools from the Internet. These perplexing assault devices and open systems have expanded the requirements for dynamic security and system security arrangements [22].

The security attack is any activity that breaches the data safety of any association utilising any procedure that it seeks to identify. It also refers to a process whereby a user compromises a computer’s security by installing malicious software, such as adware, viruses, spyware and Trojan horses. This software often deletes certain significant files,

confuses the functions of a computer, spies on online surfing behaviours and causes advertisements to appear unexpectedly on the screen. Globally, the number of attacks against data frameworks has grown consistently. Moreover, there are many assaults against the data frameworks of organisations, for example, banks and people in general [16].

2.1.1.1. Active Attack

Dynamic assaults include changing the information stream and can be ordered into four classes: disguise using stolen login IDs and passwords, bypassing the verification system or discovering security crevices in projects [23]. In session replay assault, an approved client's login data are stolen by a programmer through acquiring the session ID. Additionally, in message alteration a gate crasher modifies the bundle header locations to guide a message to an alternative goal or adjust the information on an objective machine [24].

In recent years, the data security of web applications has come under increasing attack. Web applications are the principal goal, and aggressors arrange assaults for cross-website scripting and SQL infusion. Programmers use other invasion methods in the application layer. Web application vulnerabilities can be the result of poor information legitimacy, unreliable session administration, disorderly framework settings and imperfections in working frameworks and web server programming. The best technique for limiting web application vulnerabilities is "composing secure codes." However, many organisations do not have sufficient staff or money to carry out a full code audit to prevent mistakes. Secondly, conveying web applications quickly can lead to errors. Thirdly, while items used to dissect web applications show signs of improvement, there is still a significant amount that must be done manually, and this leads to human error. Securing an organisation's web framework creates inside and outside resistance, a strategy which must incorporate different parts of IT including the web improvement, foundation, operations and security groups. If there are many assaults, clients may deny access to a network or web asset [25].

DoS: Application layer refusal of administration assaults is one of the significant dangers of web security today. The DoS assault is the most well-known assault in system security with the development of networks [20]. DoS assault location advancements incorporate system activity recognition, and bundle content discovery is exhibited. DoS

disturbs administrations by attempting to confine access to a machine as opposed to subverting the administration itself. This sort of assault focuses on either the system's transfer speed or availability. The objective is to send a flood of bundles to the system or limit access to users. The most effective DoS assaults focus on the PC system's transfer speed or network [26, 27].

DDoS: The contraction of "Distributed Denial of Service", DDoS makes online services unavailable to users. It usually interrupts or suspends the services of its hosting server temporarily. DDoS attacks can be programmed by an attacker. Divided into many computers carrying legal IP, it is not a legal user, but a zombie machine [27]. During the attack much traffic is created on servers, the website is entered and the data obtained. The fundamental objective of a DDoS assault is to bring about harm either for notoriety or for individual reasons. The DDoS attack consists of four elements: a real attacker, handlers or masters, daemon agent or zombie hosts and victim or target host [28].

HTTP Flood Attack: In this assault, a zombie machine builds a genuine TCP association with its genuine IP address and downloads the HTTP pages. In fact, the aggressor has countless zombie machines and does not unveil their IP addresses. In this assault, the solicitations of zombie machines cannot be distinguished from the demands of ordinary clients since they have an honest arrangement and are sent through typical TCP associations. The different points of a HTTP surge assault are the following: 1) immersing the data transfer capacity through outbound movement; 2) debilitating the TCP cushions; 3) devouring the CPU cycles; 4) and expanding the memory assets [29, 30].

2.1.1.2. Passive Attack

Latent assaults are carried out by listening to, or observing, transmissions. The objective of the adversary is to obtain data that are being transmitted [31]. Passive attacks are of two kinds: the release of message contents easily understood from telephone conversations, or an electronic email message, an exchanged document which may contain private data. The second sort of detached assault, activity investigation, is inconspicuous. Assume that we have a method for veiling the content of messages or other data movement so that adversaries, regardless of the possibility that they intercepted the message, could not remove the data. Aloof assaults are difficult to distinguish because they do not include any

modification of the information. Commonly, the message activity is not sent and received in a typical manner, and neither the sender nor the recipient knows that an outsider has perused the messages or watched the movement design. Be that as it may, it is conceivable to keep the achievement of these assaults, as a rule by methods for encryption. In this regard, the management of uninvolved assaults is based on aversion as opposed to location [32].

2.1.2. Security Properties

Security property tries to clarify the types of threats that exist. First, the requirements for security are defined. The significant requirements for every network security are to support the network based on the Confidentiality, Integrity and Availability (CIA) model [33]. A straightforward yet generally appropriate security model is the CIA group of three keys rules which ought to be ensured in any kind of secure framework. These standards are relevant to the whole subject of security analysis, from entry to a client's web history to the security of encoded information. It can host genuine results in the case of breaking any one of these principles [23, 24, 34].

2.1.3. Confidentiality

Secrecy is the other term used for confidentiality, which requires that data are only available to approved individuals. This includes printing, showing and uncovering the presence of a protest. Confidentiality refers to the ability to hide information from unauthorised people. The data are highly classified and kept securely from other persons who are not intended to receive them. Confidentiality also focuses on a vital component of network security. It is perhaps the clearest part of the CIA set of three with regard to security; yet, correspondingly, it can be for the most part assaulted. Secrecy is generally identical to protection. Measures are taken to guarantee privacy and to keep delicate data from reaching the wrong individuals [24].

2.1.3.1. Integrity

Integrity requires the modification of data just by approved persons. The alteration incorporates composing, changing, evolving status, erasing and making. Integrity refers to the ability to ensure that the original secure information is accurate and unchanged. Honesty means maintaining the consistency, exactness and reliability of information over its whole lifecycle. Information must not be changed, and steps must be taken to guarantee that it cannot be adjusted by unapproved individuals [25].

2.1.3.2. Availability

Availability also refers to the data that can only be accessed by authorised individuals. It is essential to make certain that the data are promptly available in all circumstances. Some types of security assaults endeavour to prevent access to suitable clients, either by hindering them or on account of auxiliary impact. The shields that are used against information interferences must take into account catastrophic events including fire. In such cases, to prevent damage to information, a duplicate might be placed in a geographically-detached area, maybe even in a flame resistant or waterproof safe. Additional security hardware, such as firewalls and intermediary servers, can be used against inaccessible information because of malicious activities; for example, refusal of administration assaults and system interruption [25, 34].

2.1.4. Security Mechanisms

Mechanisms can be nontechnical, such as asking for proof of identity before changing a password. In fact, some mechanism procedures are requested where technology cannot be enforced. Securing a system comprises a few vital steps. The initial step is to design an effective validation methodology. It is important to ensure that clients attempting to obtain data really are who they claim to be. There are different approaches to executing this verification, notwithstanding fundamental components, for example, accreditation input gadgets, prerequisites on client information and layering of information approval. This

confirmation utilises various strategies and conventions to issue endorsements and pass information bundles forwards and backwards [35].

The security systems are used to identify security assaults. A wide range of security plans can be used to counter malevolent assaults and these can be abnormal state and low-level. Security components must be adaptable to substantial systems while maintaining high calculation and correspondence proficiency. A security mechanism must be defined based on the requirement of parties to achieve high-level security for resisting attacks. The mechanism policy between parties needs to be standardised to remove the inconsistency in data transmission. Without having a security mechanism between parties, each member has to define the mechanism for security based on requirements. In such a situation, the mechanism cannot be developed and the attackers can easily figure out the mechanism, how it works and how to influence the system [26].

2.1.4.1. Firewall

Firewall refers to a computer security system used for network protection from intruders, hackers and automated programs. It consists of a software program or hardware device running on a secure host computer [29]. In such cases, it must have two network interfaces at least: one for the network intended for protection, and one for the network exposed to attack. A network firewall serves the junction point or gateway between the two networks, usually a public network and private network; for instance, the Internet. Firewalls can also be specific kinds of network traffic filters, protocol filtering and channel parcels in view of their source, port numbers and goal addresses [36].

2.1.4.2. VPN

The virtual private network (VPN) offers an appealing solution for system administrators in terms of disseminated processing. VPN consists of a set of PCs that have a generally secure system and utilise encryption and uncommon conventions. Additionally, VPN is a strategy which is used to add security and protection to private and open systems, for example, Wi-Fi hotspots and the Internet [24]. VPNs are for the most part utilised by organisations to secure delicate information. VPN has an open system to interface with

remote areas or clients. Rather than utilising a devoted, certifiable association, such as a rented line, a VPN uses "virtual" associations steered through the Internet from the organisation's private system to the remote webpage or representative. Regular VPN is Remote-get to this is a client to-LAN association utilised by an organisation, which has representatives who need to interface with the private system from various remote areas. Furthermore, site-to-webpage is done through the utilisation of committed hardware and vast scale encryption. Numerous areas can be associated with organisations over an open system such as the web. VPNs provide four basic capacities to guarantee the security of information: validation, get to control, privacy and information trustworthiness [29, 37].

2.1.4.3. UTM

UTM, which stands for Unified Threat Management, is a data security term that defines a solitary security arrangement. It is a rule of solitary security apparatus, which gives various security capacities at a solitary point in the system. A UTM machine will as a rule incorporate capacities, for example, antivirus, against spyware and spam, and arrange firewalling, interruption recognition and aversion, content separating and whole anticipation [29]. Some units, additionally, provide administration, for example, remote steering, and organise address interpretation and a virtual private system (VPN) bolster. The appeal of the arrangement depends on its straightforwardness, so that associations that had singular sellers or apparatuses for each unique security undertaking can now have them all under one merchant umbrella, supported by one IT group [38, 39].

2.1.4.4. IDS/IPS

An Intrusion Detection System (IDS) is a kind of security programming, which functions when somebody or something attempts to break data frameworks through malevolent exercises or security strategy infringement [35]. An IDS works by checking framework actions through testing vulnerabilities, as well as the documents, and directing an investigation in view of known assaults. It screens the Internet to look for the most recent dangers, which could lead to assault in the future [33, 37].

Interruption counteractive action is a pre-emptive way to deal with system security. It is utilised to recognise potential dangers and to react to them quickly. As in the interruption recognition framework, an interruption avoidance framework screens organised movement [35]. However, as changes can be made rapidly after the assailant obtains entry, interruption anticipation frameworks can additionally move promptly because of the arrangement of standards built by the system [29].

IDS and IPS systems are related to firewalls and try to search for an anomalous activity, such as an unusually large amount of traffic targeting a given host or port number, and generate alarms for network managers or perhaps even take direct action to limit a possible attack. Although this has commercial applicability, it is still a developing field [40].

2.2. Detailed Explanation of CAPTCHA Types

There are many techniques used in CAPTCHA tests, and each has its own advantages and disadvantages. Some of them are simple to implement, but easy to break. An example of a standard CAPTCHA type is shown below:

Figure 2.1. CAPTCHA.

2.2.1. Text-based CAPTCHA

With the development of technology, attackers are constantly using new and stronger techniques in order to bypass CAPTCHA tests and to enter websites. At the same time, developers are continually strengthening the shape and type of the CAPTCHA, by implementing more complicated techniques, in order to prevent every kind of attack. This is

why many different kinds of CAPTCHA exist today, and the most common is the text-based. Since the first application of the CAPTCHA test as a security method, the texting technique has been used in different ways. When the user wants to enter a website, he/she must pass the CAPTCHA test. If he/she is able to recognise the text which is shown, then he can enter. Those programs that are trained for an attack cannot succeed in this test and cannot enter the website. However, the attackers can crash these kinds of tests by using OCR. In response, developers tried to alter the tests so that even the OCR program could not succeed, leading to the development of Anti-Recognition techniques [41].

Figure 2.2. Text-based CAPTCHA [5].

More than 15 different types of text-based CAPTCHAs exist. The strongest are those techniques that are used by websites such as Google and eBay. The others are relatively easier to crash such as Baidu, NIH, Wikipedia, Blizard and Strok [12]. The shapes of these CAPTCHAs are shown in this image:

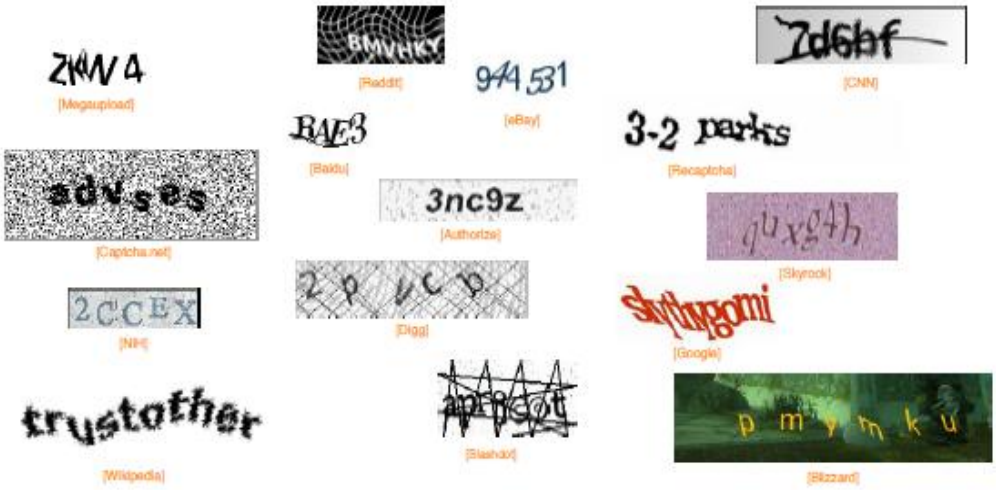

Figure 2.3. Simple CAPTCHAs [5].

Another type of text CAPTCHA uses the question and answer technique. The user should be able to answer some simple questions such as the following:

1. What is twenty one plus two?
2. What is the fourth letter in the word school?
3. If yesterday was Friday, what is today?
4. Which of the words yellow, Thursday and Richard is a colour?
5. What is six minus one (6-1=?).

Such questions are simple for humans to answer, but extremely difficult for a program to understand [4, 9]. However, by integrating natural language processing (NLP) in OCR, these types of CAPTCHAs can also be broken.

Figure 2.4. Math CAPTCHA [12].

Developers are trying to secure websites by employing different CAPTCHA techniques. Nonetheless, the attackers are constantly seeking to bypass these security measures. Nowadays, they use different machine learning or recognition systems to recognise the text CAPTCHA such as neural networks. To recognise the text, the attacker must fulfil certain stages:

1. Pre-processing: In this step, the foundation is evacuated and CAPTCHA is quality put away by the lattice.

2. Division: At this point, DeCAPTCHA tries to fragment CAPTCHAs utilising different techniques similar to shading satisfying division, which uses paint container shading surge calculation.

3. Post-division: The section's size is standardised for less demanding acknowledgment.

4. Acknowledgment: Classifiers are taught to look for letters after the division of the CAPTCHA.

5. Post-handling: The yield of the classifier makes progress [2].

Figure 2.5. Text Based on CAPTCHA

Based on assessment and comparisons, it is argued that using text in a CAPTCHA will not be an effective solution for securing websites. If it is too simple, then the attacker can easily solve the text via OCR. And if complex CAPTCHA techniques or text are used then the number of visits to the websites may decrease [15].

2.2.2. Image-based CAPTCHA

In image-based CAPTCHAs, the clients need to find identical pictures. Usually, a client is required to distinguish pictures based on similarities. The advantage of this type of CAPTCHA is that the task is a difficult AI issue compared with text recognition. Therefore, it is difficult to pass the test. In another image-based CAPTCHA, pictures are shown to the

client; every picture is unique in relation to the next. In the test, an object is given, and the client has to choose the corresponding image. An example of an image-based CAPTCHA is shown below.

Figure 2.6. Image-based CAPTCHA [7].

This kind of CAPTCHA is a successful security technique because a human user can recognise these images, but simple bots cannot understand the tests. In order to analyse and pass this kind of test, very advanced AI techniques need to be employed. Therefore, security is ensured by providing a new set of images every time the website is visited [42].

There is another image-based CAPTCHA technique which may be even more successful. In this type of CAPTCHA, an image containing an object and some text is shown to the user. Only one of the embedded texts corresponds to the correct name of the object or animal. The users have to choose and type the name in the provided field. Every time a new object is shown and embedded texts are always presented in different styles and colours. This trivial task requires a very strong OCR and object recognition technique to pass [43].

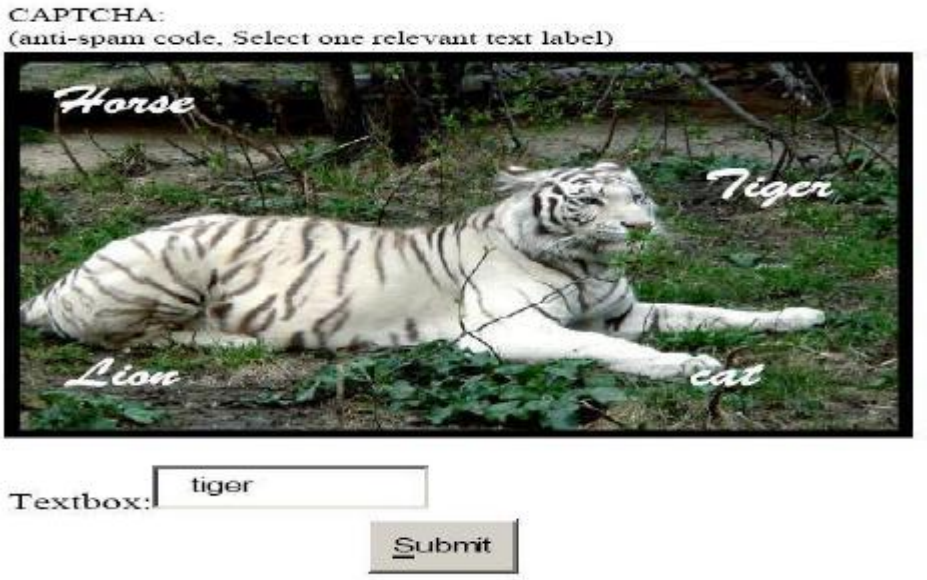

Figure 2.7. Image-based CAPTCHA [19].

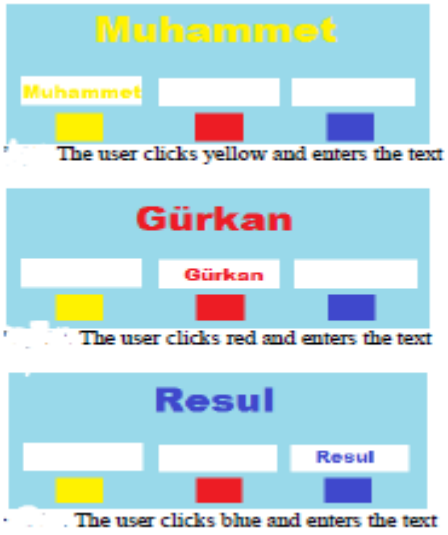

Figure 2.8. Layered text and colour combination CAPTCHA [43]

The image-based CAPTCHA can successfully resist most attackers in other ways. For example, the test may contain three pictures, each of which has more than 10 name options. The correct description of the image is also provided in those 10 options. The user has to select the correct name for the object in each picture from a corresponding combo-box in order to enter the site.

In the following figure, three pictures are shown, each of which has a drop-down menu with 10 choices [12].

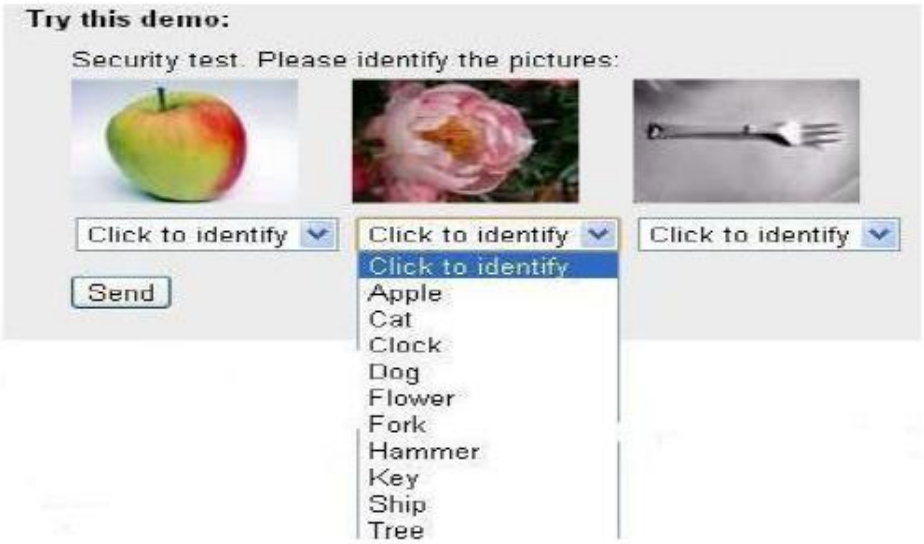

Figure 2.9. Option image CAPTCHA [19].

Using this type of test in a CAPTCHA can prove very successful, but it is very language-specific. The user must know English in order to choose the correct name for each object.

Another image-based CAPTCHA relies on image rotation. Compared with others, this is not a language-specific method but very user-friendly. The user is required to rotate a picture determined by the correct angles [15], as shown in the below image:

Figure 2.10. Rotate image CAPTCHA [9].

The developers are constantly attempting to resist attackers by using many different but simple techniques, in order not to dissuade users from visiting the websites. At the same time, they must make tests that everyone can understand, not only those who know English. Therefore, the developers select different kinds of tests [41].

Another method is based on ordering images. When a user wants to enter a website, the images pertaining to the steps of a process are shown in the test. The user is asked to arrange the pictures by placing first that which represents the first step of the process and so on. This technique is successful because it cannot be broken by the attacker. In the below image, an egg, chick and chicken are shown. The user has to put the egg in the first table, the chick in the second table and the chicken in the third table [42].

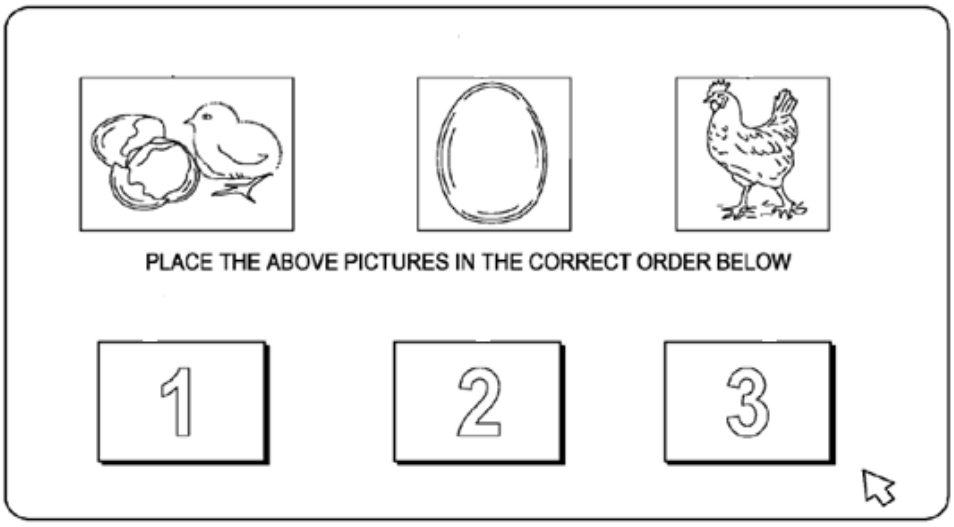

Figure 2.11. Ordering image CAPTCHA [4].

2.2.3. Video-based CAPTCHA

A video-based CAPTCHA is a recent development and less commonly seen. The user is presented with a video and requested to watch and type the most relevant words regarding the content. An example is shown below:

Figure 2.12. Video-based CAPTCHA [47].

In an enhanced video CAPTCHA for web administration enrollment, the site page requests that the client watch the video and identify what is being shown. Comprehension of the video content is demonstrated by selecting the correct options [44].

In video-based CAPTCHAs the primary hindrance is that the client's description of the video content must correspond to that contained in the information base. The other hindrance is that it consumes more memory and requires significant investment [45, 46, 47]. Of course, the different language spoken by the user and that used in the video is another problem.

2.2.4. Audio-based CAPTCHA

There are many CAPTCHA techniques utilised against attackers to protect websites. Developers are continually seeking to protect websites by using strong but simple methods. They are, for instance, using CAPTCHA sound to prevent attackers from entering websites. However, using this technique may not always be successful because of the ever-changing field of AI; for example, speech recognition systems can pass these types of tests. Language is also a big concern.

This type of CAPTCHA works in the following way:

- The user is presented with a CAPTCHA sound test.
- He listens to the sound.
- He types what is heard in a determined position.
- If the entered words are correct, then he succeeds in this test and enters the website.

If the user is a machine, it may still be successful in this test if it uses a well-trained Speech Recognition system. An example is shown in the below image [44, 48]:

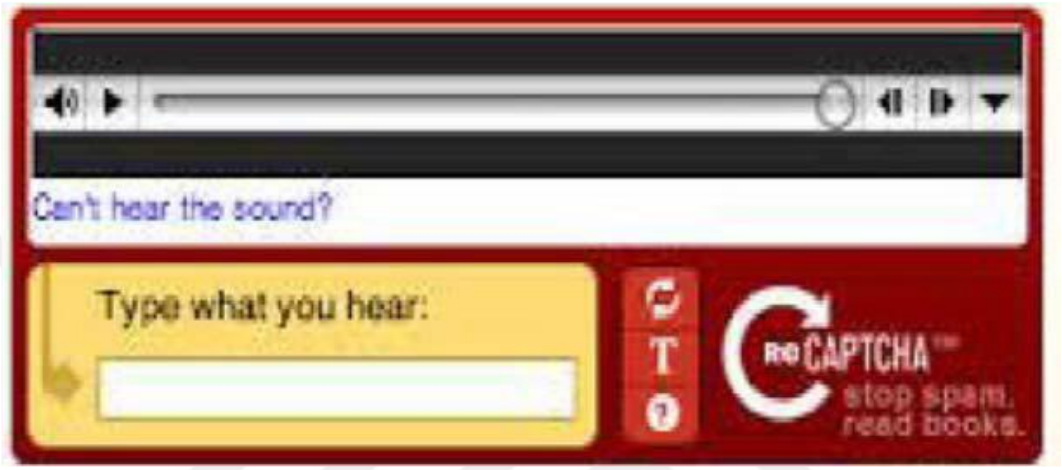

Figure 2.13. Audio-based CAPTCHA [16]

In this CAPTCHA, users, for the most part, are faced by this kind of sound test. However, if they cannot play the sound, they cannot complete the CAPTCHA [47].

2.2.5. Puzzle-based CAPTCHA

In this type of CAPTCHA, the user is asked to complete a simple puzzle in order to recompose the original image. This test requires the user to inspect the original complete picture [49, 50]. Although the test usually includes a very simple puzzle, this may still be somewhat challenging for some users to complete quickly. An example is shown below:

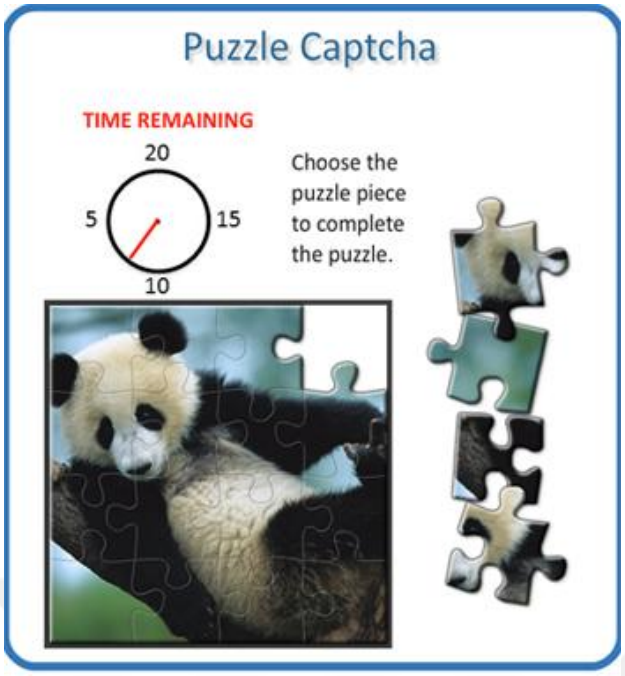

Figure 2.14. Puzzle-based CAPTCHA.

In this type of CAPTCHA, if the user does not complete the puzzle, he fails the test. Another problem is that if the bots have a product that can solve the given picture, then the web administration can be breached [50, 51].

2.2.6. Question-based CAPTCHA

This type of CAPTCHA works by requiring a user to answer a question, such as the following: "There are two pencils, three books and 1 mouse on the table. What number of organic products are there on the table altogether?" Here, the answer is 0 as there are no organic products on the table. As can be seen, the user has only to enter a number. The different qualities and inquiries can be increased by incorporating diverse examples, which adds further complexity.

However, the whole question is not necessarily a content configuration. In the setup, pictures of words, for instance, can be included, such as organic and natural products. This is illustrated in the following example:

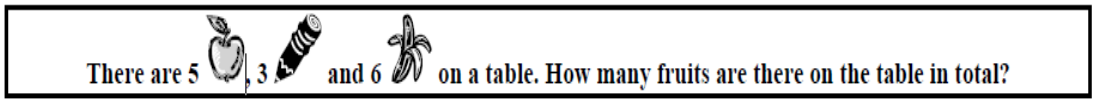

Figure 2.15. Choose the correct answer

The correct answer to the question posed is 11 fruits .The PC may be able to separate content from pictures, but this would be a difficult task. The likelihood of effectively answering this sort of question is low considering that the PC requires accompanying capacities:

1- The computer must see the pictures through OCR-based programming.
2- The computer must perceive shapes in the picture. Obviously, it must separate text from image, which is a difficult task without human input.
3- After distinguishing text from pictures, a PC must have the capacity to understand the question.
4- Even if a PC negotiates each of the mentioned stages effectively, it must be fundamentally equipped to note the indicated address [51].

Figure 2.16. Question-based CAPTCHA.

This procedure, which is unlike the OCR-based CAPTCHA systems, requires simply giving a number as the correct answer. Therefore, it is not difficult for users and does not take up too much time. In this method, entering a number is all that is required. Therefore, it can be used on devices which do not have support, for instance, mobile phones and Pocket PCs. This procedure can be executed on little devices and devices with limited resources [51].

Colour image-based CAPTCHA. In this type of CAPTCHA, the user is required to describe a certain picture. For instance, the picture of a car is accompanied by the following statement: “Enter the colour of the car.”

Figure 2.17. Colour image-based CAPTCHA.

Multi-colour image-based CAPTCHA. In a multicoloured CAPTCHA, the test can ask different questions and make requests.

- How many colours are there is the picture?
- Enter the name of the colour in the centre.
- Enter the colour of the most extreme percentile in the picture.
- Enter the colour on the left side or the right side.

Figure 2.18. Multi-colour image-based CAPTCHA.

Bongo-based CAPTCHA. Another example of a CAPTCHA is the BONGO program. The user is presented with two sets of images. The sets differ in one specific attribute, which must be recognised by the user to pass the test. The properties are, for example, the thickness of the lines, the different shapes, the size of the pictures. Once the user has recognised the particular attribute, he has to add some images to one of the two sets [41].

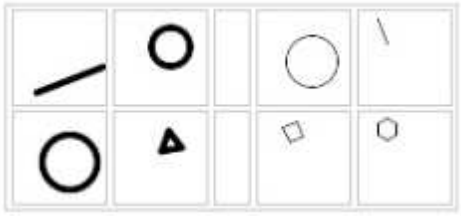

Figure 2.19. Bongo CAPTCHA [16].

These two sets are unmistakable as the shapes on the left are drawn with thick lines and everything on the right is drawn with faint lines. The user has to choose the right course of action; of the four single squares, he is asked to select where each piece fits. The user can easily pass the test if he chooses the correct set of squares. It is important to ensure that a user is not confused by too many choices [41].

2.3. CAPTCHA Drawbacks

Table 2.1. CAPTCHA Drawbacks [4, 14, 19].

Types of CAPTCHA	Strength point	Weakness point
Text based CAPTCHA	• Relaxed implementation • Simple use/User friendly	1. Complex text: the user may have some problems to identify the correct text or characters. 2. It can be easily recognized by OCR techniques.
Image based CAPTCHA	• Hard to break, • Simple use/User friendly	Some users face problems with image identification due to low vision or blurred of images.
Video based CAPTCHA	• Harder to break	1. Not practical due to the large size of the file. 2. Not user friendly/annoys users
Audio based CAPTCHA	• Requires sophisticated methods to break, like speech recognition.	1. Addresses only English speakers 2. The character may have a similar sound.
Puzzle based captcha	• It seems like a fun. • It helps the user to monitor their brain. • It's like a game so the user can more interact with this captcha system.	1. Time consuming. 2. User cannot identify the puzzle easily/Not user friendly

2.4. A CAPTCHA Recognition Approach

In this section, the steps for a CAPTCHA recognition strategy are proposed [52].

2.4.1. Pre-processing

In order to achieve the main objective of the CAPTCHA, the images of the test should be contorted and altered so that they cannot be interpreted by OCR techniques easily. Most importantly, noise should be included in the picture, especially in the background, to make the automated extraction of the text difficult. Therefore, the first step in CAPTCHA recognition in [53] is pre-processing to eliminate such noise from the extraction text. After converting the image to binary via Otsu's method, due to its relatively higher performance compared with others, the adaptive median filter is used to eliminate noise [54, 55].

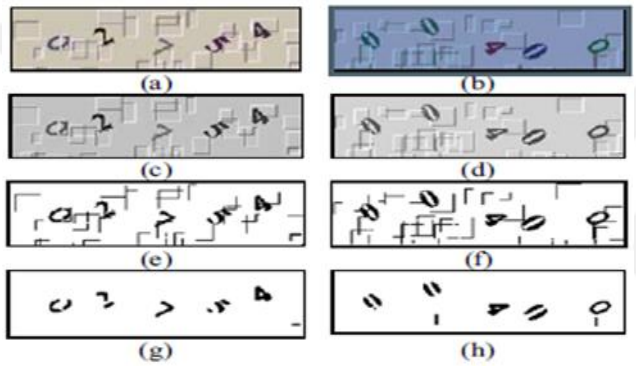

Figure 2.20. a-b (original images), c-d (grey-level images), e-f (binary images), g-h (after noise reduction) [52].

2.4.2. Number Splitting Procedure

After the pre-processing, the next step is to split the numbers. To do this, a projection method is employed. The black pixels are projected onto the x-axis and the coordinates of each digit (0-9) are determined in the CAPTCHA image. It should be noted that the projection of multiple digits may overlap in some cases [56].

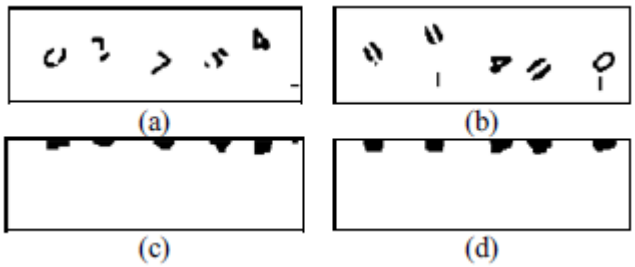

Figure 2.21. Split the numbers

2.4.3. The probability Pattern Module

In the next step, the tilted numbers need to be considered. The digits in a CAPTCHA image may be tilted from -45 to 45 degrees. In order to recognise tilted digits properly, a probability pattern module is introduced in [52] with the following equations. They propose that the digits tilted from 0 to 45 degrees should be handled separately from those tilted from -45 to 0 degrees. They also provide training dataset examples for recognising such digits as shown in Figure 2.22.

$$W(x,y) = \frac{\sum hk(x,y)}{count(k) * zi} \quad (1)$$

hk (x ,y) = 1 if (x ,y) is the pixel of target number or 0 otherwise (2)

$$\mathbf{Z}i = \sum \frac{\sum hk(x,y)}{count(k)} \quad (3)$$

The Wi (x, y) is a weight capacity of pixel located probability for every objective number (i = 0 to 9); where (x, y) is the imaging coordinate. hk (x, y) is the theory capacity of preparing set (k=0 to 9) and the Count (k) is the quantity of preparing set about k. On the off chance that (x, y) is the pixel of the target number, then hk(x, y) is equivalent to 1. Zi is a standardised capacity.

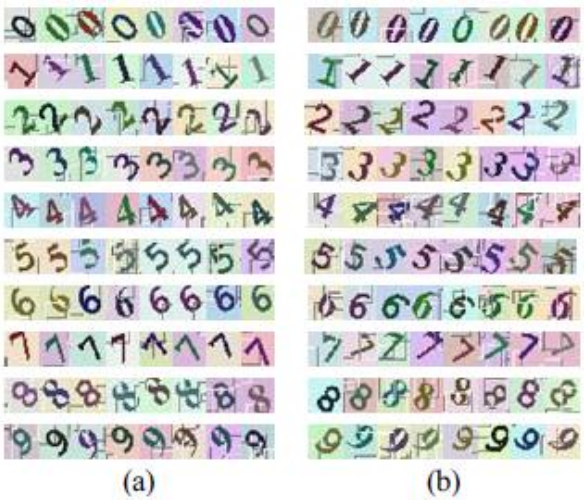

Figure 2.22. a (-45 to 0 degree tilted) and b (0 to 45 degree tilted) [52].

2.4.4. Number Recognition CAPTCHA Image

As a last step, score matching is used to determine each split digit. The data set in Figure 2.22 determines the similarity score for each digit using the following equation. Whatever maximises this similarity score becomes the final answer for that digit. After finding the final answer for all digits, the CAPTHCA can be read successfully.

$$\text{Matric} = \max_{0 \le i \le 9} \left(\sum \text{Wi}\,(x, y) \times \text{hi}(x, y) \right) \quad (4)$$

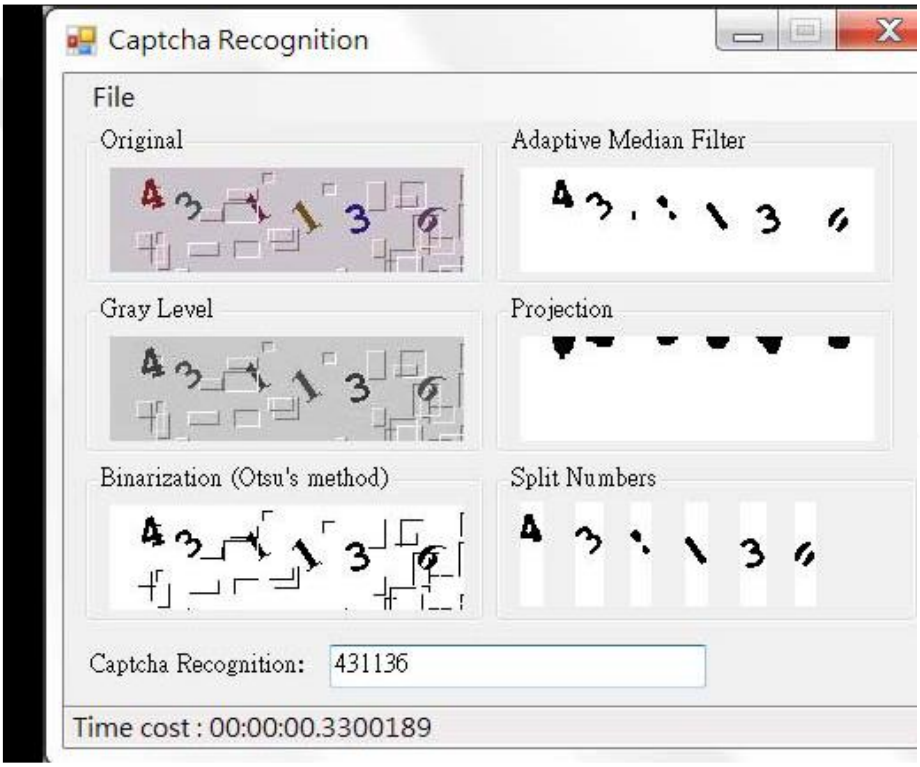

Figure 2.23. The sample of suggested CAPTCHA recognition [52].

2.5. Breaking CAPTCHAs

As previously mentioned, attacks on CAPTCHAs are increasing, and malicious users try to break even the most advanced CAPTCHA tests. In response, developers build stronger CAPTCHAs to ensure the safety of their websites. One framework for breaking a

CAPTCHA is explained in section 3. There are similar studies that investigate the weakness of CAPTCHAs to attacks. One such study presents several examples of attacks on widely used website CAPTCHAs such as Yahoo! in order to evaluate their robustness [45, 57].

Like most other object or pattern recognition strategies, the case studies presented here also include three main steps: (1) pre-processing, (2) extraction and segmentation and (3) extraction. In the pre-processing stage, in order to eliminate noise in the Yahoo! CATCHPAs, a threshold method is used to convert all pixels with higher intensity than the threshold. These are converted to white and others are converted to black. The following figure shows this pre-processing step [58, 59].

The second step, extraction and segmentation, includes two sub-steps to determine the location of characters. To extract characters several methods are used such as cut head and tail (CHT), guide lines, loop principle, projection with guide lines and even cut. The appropriate one is chosen and applied depending on the CAPTCHA [58].

- Cut head and tail (CHT). This method is used to detect the first and last characters since they have only one neighbouring character. Therefore, only the horizontal projection together with area colour filling would suffice to determine the two characters by analysing their exposed sides. Note that certain characters may have similar appearance after the CHT and area colour filling. Some examples of characters that have similar and different appearances are shown in the following two figures.

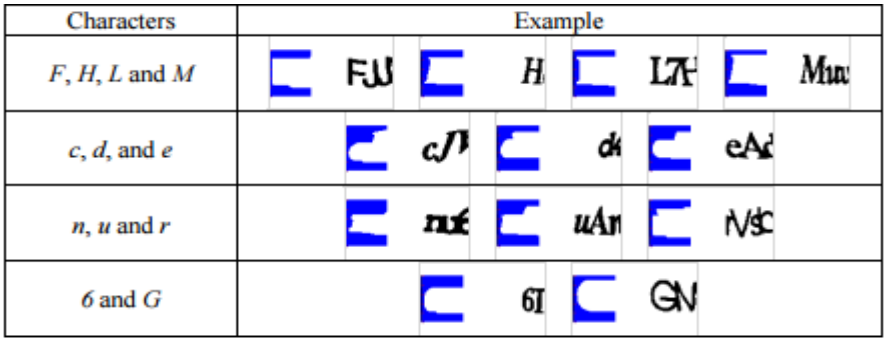

Figure 2.24. Characters with same or similar projection features.

- Guide lines principle. In order to recognise characters, in this technique the characters are separated into their bottom, middle and top regions. The Yahoo! CAPTCHA follows this guideline principle, which can be used to break the CAPTCHA test. The following figure shows the base and guide lines.

Figure 2.25. Guide lines principle

• Loop principle. This technique is used to determine whether a loop belongs to the characters. For example, digit 8 includes two loops and they determine that the two loops are parts of digit 8, if they are aligned vertically and do not overlap. Some characters when written connected, may create additional loops in between.

• Projection with guide lines: This is another technique used to recognise characters. The upper and lower characters separated by the middle layer are projected with respect to the mean and base line, respectively. The projections provide information about the categorisation of the characters. An example projection is shown in the following figure (Detect G and L (Left), Detect H (Right)).

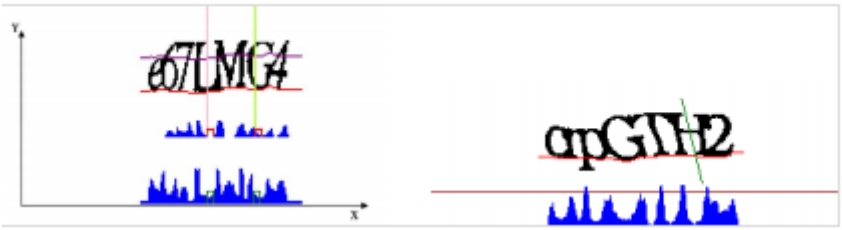

Figure 2.26. Detect G and L (Left), Detect H (Right).

• Even cut method. After the segmentation, there may still be more than one character in a chunk. Counting the number of pixels in chunks, it is determined whether there is more than one character in the chunk, to split or segment the chunk further.

Finally, in the extraction step, the features for each character are analysed in the Yahoo! CAPTCHA and recognition is made in this way.

Following similar steps in [58], a comprehensive step-by-step execution of attacks on Yahoo!, Baidu and reCAPTCHA is explained, and the success or accuracy of each is reported and compared with the Stanford attack.

3. PROPOSED TECHNIQUES

In this chapter, the proposed approach for analysing the question-based CAPTCHA and implementation steps are discussed.

3.1. Website Prototyping

For implementing the proposed technique, an online user interface has been created (http://www.bcaptcha.com). As this research investigates breaking the question-based CAPTCHAs, it is necessary to have a prototype of the CAPTCHA mechanism of some Turkish universities that protect student information. Working directly on the main website of the university and trying to break the question-based CAPTCHA is considered as a cyber crime and is copyrighted to the website.

3.1.1. Question-based CAPTCHA Generation

In the prototyped domain (http://www.bcaptcha.com), all the required settings and the user interface are copied from the Adiyaman University and then some of the CAPTCHA images were obtained from the CAPTCHA system of the Adiyaman University website. Then we used their content to design our website because we did not want work on the University's server.

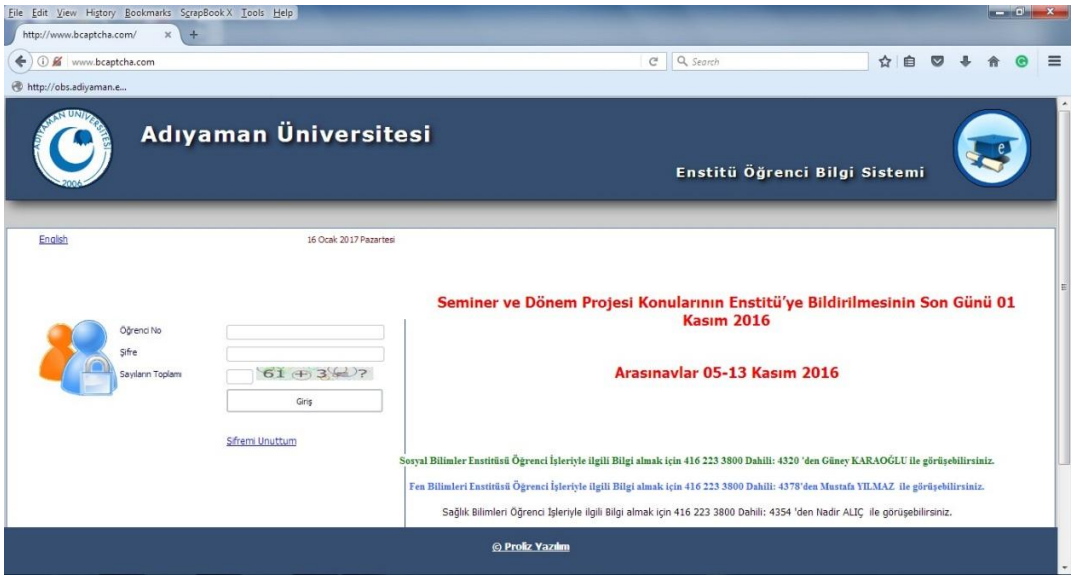

Figure 3.1. The prototyped website (http://www.bcaptcha.com).

In the main website, 53 sample question-based CAPTCHAs are uploaded. By using the PHP code, every time the user reloads the page a new image for a question-based CAPTCHA is shown.

3.1.2. BCAPTCHA Database

To register any user, a username and password are needed. A database was built in MySQL, which contains some usernames and passwords. This makes it possible for the user to login successfully to the system by entering a correct username/password and CAPTCHA. The system will then redirect the user to the main page if all the fields are completed correctly.

3.2. Browser Add-on

In this study, an add-on is used to save the presented CAPTCHA image in the browser and then make this an input for processing. The Scrapbook add-on, which is an extension for Mozilla Firefox browser, allows the user to automate the process of saving the image to the local computer. The Scrapbook add-on saves images in the webpage and easily manages collections. The key features are lightness, speed, accuracy and multi-language support.

It helps you except? Web pages and organize the collection.

This add-on is based on ScrapBook Plus and also integrated several features in ScrapBook.

1. Saving the web page faithfully: Web pages shown on the screen can be saved without losing any subtle detail. The saving time and source URL are also recorded for later reference.
2. Saving partial content: You can save partial web content, save without images and videos, save without styles, save without scripts, edit before save, and more ways for saving are available for you.
3. Extensive saving: You can save web pages and files linked by the web page, save multiple opened tabs, save a list of pages using a URL list, and more batch save functionalities are available for you.

4. Manage data: You can manage saved items with a tree structure, just as easy as managing the bookmarks.
6. Searching data: You can search any fragment of the saved web pages with the built-in full-text engine.
7. Editing data: You can add highlights, comments, annotations, or even edit the source HTML for the saved pages.
8. Taking notes: You can create note pages in ScrapBook, and edit them as easy as editing web pages.
9. Input and output data: You can combine multiple data items into one. You can generate HTML tree list and make a static scrapbook site. You can configure a multi-ScrapBook database which won't interfere with each other. You can import and export data items for backup or exchange.
10. Add-ons: Some Firefox add-ons can be integrated with ScrapBook and extends its power, such as these ones.

Installation

Just download and install the desired version here or from the history list. If you were using ScrapBook or ScrapBook Plus, be sure to disable or remove them before using ScrapBook X, to prevent a possible conflict.

3.3. The CAPTCHA Recognition by Proposed OCR System

After downloading the image by Scrapbook add-on, the original image is processed by preparing for recognition and determining the equation used in the image.

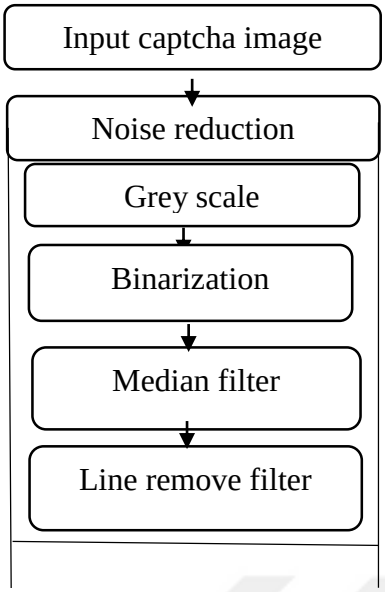

Figure 3.2. Pre-processing noise reduction flow chart diagram.

A software (BCAPTCHA) was built for automatically saving the image by using add-on and preparing the image for pre-processing and then recognition. In the first step, the software uses a keyword to make the image ready for processing. The BCATCHA software, which has been written in Java programming language, uses the robot object to press the (c) key on the keyboard to save the loaded image in the browser.

3.3.1. Noise Reduction

The CAPTCHA pictures have a great deal of commotion. Consequently, a pre-processing system is connected to kill the commotion. The pre-handling system is shown in Figure 3.3. The basic picture handling is utilised to upgrade the power of the source CAPTCHA pictures. To begin with, the CAPTCHA picture is changed to a grayscale picture. Then, the commotion is expelled using adaptive median filter (AMF). The AMF beats all current middle based channels that can lessen commotion.

With the end goal of planning and improving question-based CAPTCHAs, a few unique capacities and methods are connected; for example, differentiating pictures, changing question-based CAPTCHA pictures from shading mode to grayscale mode, thresholding CAPTCHA pictures, which is the most critical system that binarises question-based CAPTCHAs, reducing noise by expelling little dots on the pictures, and playing out some

morphological operations on the grounds that through actualising different pre-processing forms the results may contain gaps in a few sections of the pictures, which may evacuate some pixels in the middle channel of digits on pictures that have ordinary numbers for questions.

Figure 3.3. The image before processing.

Figure 3.4. The image after noise reduction.

As shown in figure 3.4, after applying noise reduction to the original image, this makes the image ready for OCR CAPTCHA recognition.

3.3.2. Grayscale

After applying contrast work to the CAPTCHA images, in the second critical picture pre-processing system a grayscale work is applied to pictures for the questions. One of the most essential and straightforward strategies of picture upgrade procedures is grayscale transformation. Diminishing computational necessities and rearranging binarisation is the principle reason for applying grayscale change to images, rather than working on shading pictures of the question-based CAPTCHAs, specifically in binarisation handle.

To change pictures from shading mode to dark mode, a technique or capacity from Image Magic library was utilised (- shading space dim). Figure 3.5 demonstrates the transformation of a differentiated CAPTCHA picture to a dark scale picture. The differentiated picture has some unique shading. However, when applying grayscale change, the foundation shading is obviously adjusted. In light of our trials, such change is one of the principle components for better application of binarisation procedure with less computational prerequisites.

Figure 3.5. Applying thresholding technique to a contrasted and grayscale image.

3.3.3. Binarisation

Isolating picture pixels as background (White) or text (Black) is the procedure of binarisation, which is incorporated into OCR frameworks as a pre-processing method. Isolating pixels is the most critical process that empowers the division process and character acknowledgment procedure to identify and break down the areas of the picture which probably contain text.

Binarisation is performed using thresholding. In a picture, isolating the data, such as content from its experience, is the procedure of image thresholding. This is performed when dealing with pictures of a shading record or of dim level. Worldwide, thresholding and nearby thresholding are the most popular. At the point of the whole archive picture, just a single edge esteem is utilised, and the global thresholding strategy is utilised, which is for the most part from the force histogram of the picture contingent on the estimation of the foundation level. However, as indicated by the data, when every pixel diverse esteem is used, the technique utilised is nearby versatile thresholding.

In this study, utilising thresholding strategy has many focal points that require applying to CAPTCHA pictures; for example, better performing division handle, acknowledgment preparation and resolving the issues of uneven lighting. However, ordinarily, photographs of CAPTCHAs have less states of uneven lighting since capacity was offered to the clients to see pictures legitimately using better lighting. Since utilising versatile thresholding creates more opportunities to process CAPTCHA pictures than worldwide thresholding, worldwide thresholding technique was adopted for binarising CAPTCHA pictures [60, 61].

To benefit from the utilisation of worldwide thresholding, (- threshold value %) was utilised. This capacity applies simultaneous dark/white limit to the picture. In view of the tests adopted, the best edge is 40%, which works really well under various conditions.

Figure 3.5 shows two CAPTCHA pictures. The primary picture is pre-handled by applying strategies of complexity and grayscale, and the second picture is binarised and delivered after applying thresholding (- threshold 40%) to the principal picture.

3.3.4. Median Filter

The vigour and accuracy of the element extraction and characterisation period of the OCR structure can be achieved by diminishing the commotion in the picture in the pre-preparing stage. There are a few strategies which are used for expelling commotion from pictures; for example, middle channels or Gaussian obscure channels. Such strategies can adequately evacuate detached pixel commotion. In this study, middle channel procedure was utilised with the goal of evacuating dots or commotions in binarised pictures. The middle channel is a unique type of low-pass channel. Middle channel strategy takes a range of pictures (width 40 pixels, high 177 pixels) and then sorts pixel esteem reduced to most noteworthy esteem. Subsequently, it extracts the middle incentive from the sorted esteem. Finally, it replaces the inside pixel estimation of the handled range with the medium esteem [63].

With the end goal of applying a middle channel system to CAPTCHA pictures, (- middle radius) was utilised. In our trials, middle channel performs better on pictures when its range is (3). Figure 3.6 indicates the impacts of applying a middle filter (- middle 3) to a CAPTCHA picture. Unmistakably, middle filter expelled the greater part of the little dots around the text.

Figure 3.6. Effects of applying median filter technique to a noisy image.

3.4. Implementing OCR Process

The image after noise reduction is sent to an OCR machine for recognition. The software processes the results of the OCR to calculate and find the final result of the question-based CAPTCHA.

As shown in the image below, the output of OCR, which is 362, is split into digits. The first two digits are combined as separate numbers (36) and the third digit as another number (2). Finally, two numbers (36 and 2) are added, and the summation is used in the software to break the question-based CAPTCHA.

Figure 3.7. Image noise reduction and its OCR-produced equation.

Figure 3.8. Run-time process.

The framework plays out a few procedures; for example, taking an image of CAPTCHAs or browsing to an image, later performing document design investigation

algorithm for identifying CAPTCHA locale from picture foundation and sending the picture to the managed record server (bcaptcha.com) for further handling. Notably, the programming framework performs the proposed procedures, which are picture pre-processing calculations, perceiving digits (0-9) by utilising Tesseract OCR mechanism for question-based CAPTCHAs, distinguishing and removing picture information in perceived digits and, finally, storing valuable data in the database. The usage details of image handling are provided in the accompanying subsections.

3.5. Dictionary Attack

A dictionary attack is an attempt to gain illicit access to a computer system by using a very large set of words to generate potential passwords. In the approach of the present study, a dictionary file is built, which contains some usernames and passwords. The usernames and passwords are used during the attack time to login into the main page of the website.

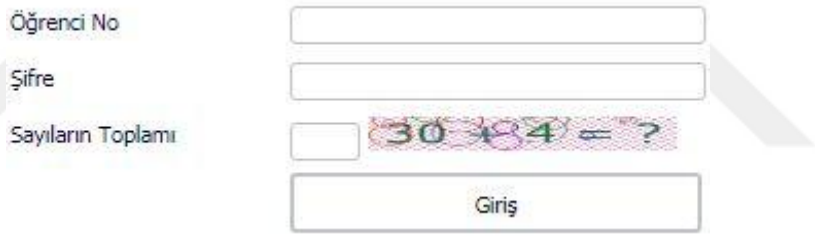

Figure 3.9. Login page during software attack: entering username and password for the correct CAPTCHA.

Figure 3.9 shows the login page of bcaptcha.com. The username and password have been extracted from the dictionary and the image CAPTCHA is recognised by OCR.

4. EVALUATION AND RESULTS

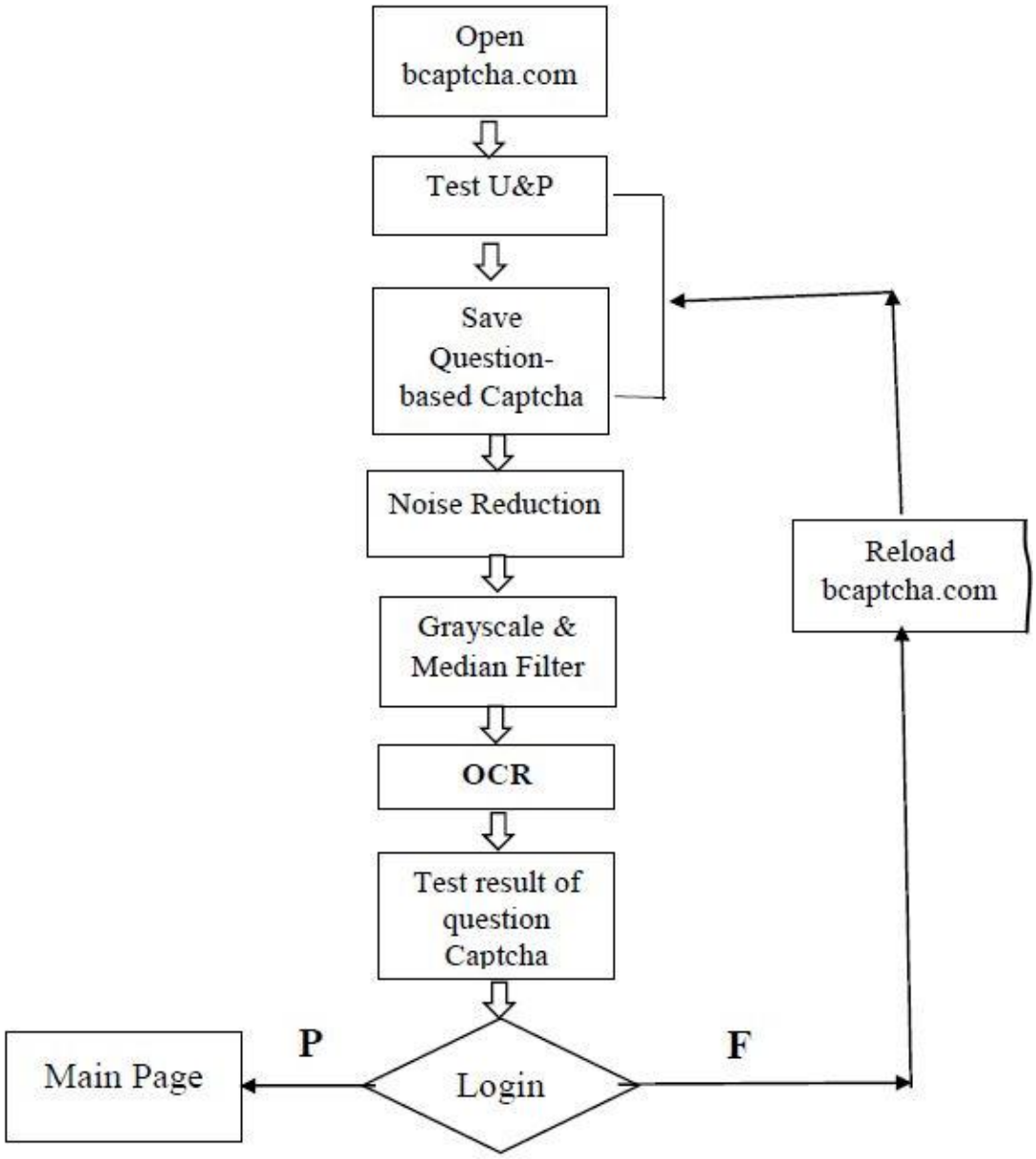

Figure 4.1. The diagram shows a system of recognition framework for question-based CAPTCHAs

4.1. Evaluation

After the detailed study of the different types of CAPTCHAs of some reliable websites and scientific resources, a comparison and evaluation was made between the different types that have been built for some Turkish universities for their student information pages. A

CAPTCHA is used to protect webpages in order to prevent being attacked and broken or crashed. In particular, the question-based CAPTCHA is a combination of two other types of CAPTCHA. These two types are text-based and image-based.

In this study, it is proved that this type of CAPTCHA can be broken and noise reduction background removed by creating software. It can also be recognised by OCR mechanisms. This break of CAPTCHA using software is caused by some weak points of technique, which are the following:

1- All the techniques of the images to produce this CAPTCHA have similar backgrounds that help the attackers to break it more easily.
2- Lack of skew and ambiguity in writing the formula/equation.
3- There is only one type of equation and text which is repeated every time (in every circumstance) the page is reloaded. The equation is written in this pattern (x + y) (38 + 2). This continuous repetition of the equation makes it easier for the hackers to break it. The equation could have been written in a different format (89 plus five, fifteen – 8 or twenty-six ÷ 2) using different signs (plus, minus, multiplication, etc.) so to make the CAPTCHA stronger and more complex.

According to the evaluation and analysis of different CAPTCHAs, the best type for protecting webpages and for users is the reCAPTCHA, which can only be passed by clicking on the checkbox, as shown in figure 4.2, one or two times. This type of CAPTCHA cannot be passed more than two times in one session. The stage will be changed if the user is suspected of being a robot.

Figure 4.2. Clicking on the check box pass of the reCAPTCHA.

Due to this suspicion, the CAPTCHA presents some complex images, which are shown in figure 4.3. The user should choose the right images to go with the above text. After

choosing the correct pictures, the user is allowed to pass. Because of this complicated equation, the machine cannot understand and identify the CAPTCHA easily.

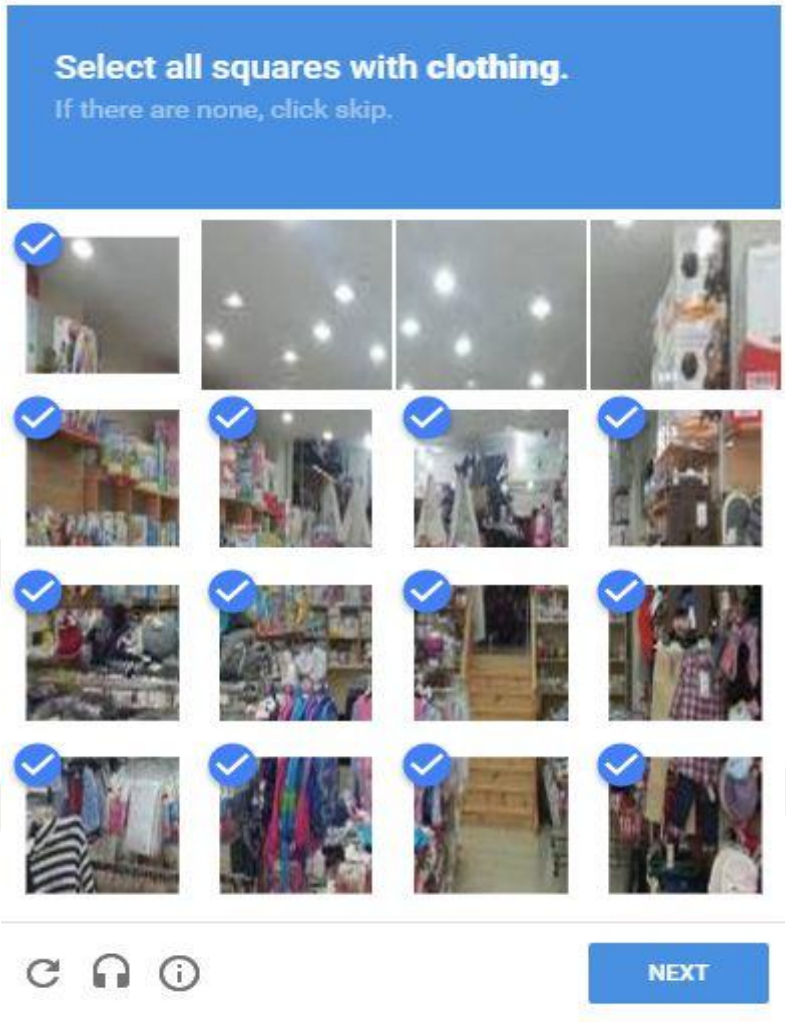

Figure 4.3. Google new reCAPTCHA.

4.2. Results

In this study, 53 CAPTCHA cases were examined. Three numbers were hidden in each CAPTCHA picture. Fifty-three cases were procured haphazardly to make the product framework, as can be seen in Figure 3.3. Figure 4.1 illustrates the CAPTCHA acknowledgment structure. Since we need to perceive the number viably, some picture handling techniques were utilised to part the numbers in the CAPTCHA pictures. We utilised match condition. The exploratory outcome appears in Figure 3.4. In the analysis, the

quantitative appraisal demonstrates that the proposed acknowledgment technique completes 92.45% of the 53 CAPTCHA images.

The software system attacks every 32 seconds for each username and passwords, and successfully recognises 49 out of 53 images in the question-based CAPTCHA.

Based on the data of the program test, it can be concluded that this type of CAPTCHA, which has been used by some Turkish universities to protect the student information system, has some weak points, which makes it unstable.

5. CONCLUSIONS AND RECOMMENDATIONS

The different types and techniques of CAPTCHA were extensively reviewed. The five most commonly used CAPTCHAs were analysed and their strengths and weaknesses evaluated. It was concluded that the text-based CAPTCHA is the most preferred, although it is more prone to being bypassed by bots.

Following this, the study presented an overview of CAPTCHA and where it is used. This included discussion of techniques used to break CAPTCHAs. The main steps for breaking a CAPTCHA were also considered, and examples of CAPTCHA attacks were provided.

The security of the CAPTCHA framework was examined as regards data framework organisation in some Turkish universities. There is much focus on the testing of CAPTCHA pictures, and a few target numbers are covered or detached by the commotion. Keeping in mind the end goal to examine these CAPTCHA pictures, image noise reduction and OCR recognition were proposed to perceive the objective numbers in all CAPTCHA cases. As the trial results show, the danger of the CAPTCHA framework as regards the data framework organisation might exist because of the proposed CAPTCHA acknowledgment strategy. It is recommended that Turkish universities strengthen their CAPTCHA framework.

A new BCAPTCHA system was presented that tries to save the images which are loaded. The experimental result is shown in figure 4.1. The quantitative evaluation demonstrates that the proposed acknowledgment strategy completes 92.45% of the 53 CAPTCHA pictures. The software system attacks every 32 seconds for each username and passwords, and successfully recognises 49 out of 53 images in the question-based CAPTCHA.

It is recommended that Turkish universities change their CAPTCHA system to reCAPTCHA, which is a free service that shields sites from spam and misuses. reCAPTCHA uses a moved possibility examination engine and adaptable CAPTCHAs to prevent robotised programming from partaking in harmful activities on a site. It does this while allowing users to enter effortlessly.

6. REFERENCES

[1] **von Ahn, L**., 2000. The official CAPTCHA site. http://www. captcha.net/.

[2] **Bursztein, E., Martin, M. and Mitchell, J.,** 2011, October. Text-based CAPTCHA strengths and weaknesses. In Proceedings of the 18th ACM conference on Computer and communications security (pp. 125-138). ACM.

[3] **Wüest, C.,** 2010. The risks of social networking. Symantec Corporation.

[4] **Choudhary, S., Saroha, R. and Dahiya, Y**., 2015. Big data: challenges and prospects. In International Conference on Recent Trends in Engineering Science and Management (pp. 821-898).

[5] **Shirali-Shahreza, M. and Shirali-Shahreza, S.,** 2007, December. Question-based CAPTCHA. In Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on (Vol. 4, pp. 54-58). IEEE.

[6] **Von Ahn, L., Blum, M., Hopper, N.J. and Langford, J.,** 2003, May. CAPTCHA: Using hard AI problems for security. In International Conference on the Theory and Applications of Cryptographic Techniques (pp. 294-311). Springer, Berlin, Heidelberg.

[7] **Zhu, B. B., Yan, J., Bao, G., Yang, M., & Xu, N.** 2014. Captcha as Graphical Passwords—A New Security Primitive Based on Hard AI Problems. Information Forensics and Security, IEEE Transactions on, 9(6), 891-904.

[8] **Tam, J., Simsa, J., Hyde, S. and Ahn, L.V.,** 2009. Breaking audio captchas. In Advances in Neural Information Processing Systems (pp. 1625-1632).

[9] **Pawar, S.E. and Bauskar Makarand, M.,** 2013. CAPTCHA: A SECURITY MEASURE AGAINST SPAM ATTACKS. no. May, pp.854-857.

[10] **Kiranjot, K, Sunny, B.** 2014. CAPTCHA and Its Techniques: A Review, International Journal of Computer Science and Information Technologies, 5, 6341-6344.

[11] **Elie, B,. Matthieu, M,. John, C. M,.** 2011. Text-based CAPTCHA Strengths and Weaknesses, ACM, 125-138.

[12] **ur Rizwan, R.,** 2012. Survey on captcha systems. Journal of Global Research in Computer Science, 3(6), pp.54-58.

[13] **Hernandez-Castro, C.J. and Ribagorda, A.,** 2010. Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study. computers & security, 29(1), pp.141-157.

[14] **Shirali-Shahreza, M.H. and Shirali-Shahreza, M.,** 2007, October. Multilingual captcha. In Computational Cybernetics, 2007. ICCC 2007. IEEE International Conference on (pp. 135-139). IEEE.

[15] **Sharma, S. and Seth, N.,** 2015. Survey of Text CAPTCHA Techniques and Attacks. International Journal of Engineering Trends and Technology (IJETT), 22(6).

[16] **Chandavale, A.A. and Sapkal, A.M.,** 2010, November. Algorithm for secured online authentication using CAPTCHA. In Emerging Trends in Engineering and Technology (ICETET), 2010 3rd International Conference on (pp. 292-297). IEEE.

[17] **Choudhary, S., Saroha, R., Dahiya, Y. and Choudhary, S.,** 2013. understanding CAPTCHA: Text and Audio Based Captcha with its Applications. International Journal of Advanced Research in Computer Science and Software Engineering, 3(6).

[18] **Fisk, F.C., Ramanathan, S., Terry, M.A. and Trevathan, M.B., International Business Machines Corporation**, 2014. Advanced CAPTCHA using images in sequence. U.S. Patent 8,713,703.

[19] **Bhalani, S.D. and Mishra, S.,** 2015. A survey on CAPTCHA technique based on drag and drop mouse action. International Journal of Technical Research and Applications, 3(2), pp.188-189.

[20] **Douligeris, C. and Mitrokotsa, A.,** 2004. DDoS attacks and defense mechanisms: classification and state-of-the-art. Computer Networks, 44(5), pp.643-666.

[21] **Saini, B.S. and Bala, A.,** 2013. A review of bot protection using CAPTCHA for web security. IOSR Journal of Computer Engineering, 6, pp.36-42.

[22] **Frias-Martinez, V., Stolfo, S.J. and Keromytis, A.D.,** 2008, December. Behavior-profile clustering for false alert reduction in anomaly detection sensors. In Computer Security Applications Conference, 2008. ACSAC 2008. Annual (pp. 367-376). IEEE.

[23] **Lano, K., Clark, D. and Androutsopoulos, K.,** 2002, September. Safety and security analysis of object-oriented models. In International Conference on Computer Safety, Reliability, and Security (pp. 82-93). Springer Berlin Heidelberg.

[24] **Schneider, S.,** 1996, May. Security properties and CSP. In Security and Privacy, 1996. Proceedings., 1996 IEEE Symposium on (pp. 174-187). IEEE.

[25] **Bowers, K.D., Juels, A. and Oprea, A.,** 2009, November. HAIL: A high-availability and integrity layer for cloud storage. In Proceedings of the 16th ACM conference on Computer and communications security (pp. 187-198). ACM..

[26] **Al-Aqrabi, H., Liu, L., Xu, J., Hill, R., Antonopoulos, N. and Zhan, Y.,** 2012, April. Investigation of IT security and compliance challenges in Security-as-a-Service for Cloud Computing. In Object/Component/Service-Oriented Real-Time Distributed Computing Workshops (ISORCW), 2012 15th IEEE International Symposium on (pp. 124-129). IEEE.

[27] **Kandula, S., Katabi, D., Jacob, M. and Berger, A.,** 2005, May. Botz-4-sale: Surviving organized DDoS attacks that mimic flash crowds. In Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2 (pp. 287-300). USENIX Association.

[28] **Feinstein, L., Schnackenberg, D., Balupari, R. and Kindred, D.,** 2003, April. Statistical approaches to DDoS attack detection and response. In DARPA Information Survivability Conference and Exposition, 2003. Proceedings (Vol. 1, pp. 303-314). IEEE.

[29] **Deng, F., Luo, A., Zhang, Y., Chen, Z., Peng, X., Jiang, X. and Peng, D.,** 2008, November. TNC-UTM: A holistic solution to secure enterprise networks. In Young Computer Scientists, 2008. ICYCS 2008. The 9th International Conference for (pp. 2240-2245). IEEE.

[30] **Yatagai, T., Isohara, T. and Sasase, I.,** 2007, August. Detection of HTTP-GET flood attack based on analysis of page access behavior. In Communications, Computers and Signal Processing, 2007. PacRim 2007. IEEE Pacific Rim Conference on (pp. 232-235). IEEE.

[31] **Schneider, D.,** 2012. The state of network security. Network Security, 2012(2), pp.14-20.

[32] **Padmavathi, D.G. and Shanmugapriya, M.,** 2009. A survey of attacks, security mechanisms and challenges in wireless sensor networks. arXiv preprint arXiv:0909.0576.

[33] **Yoshioka, N., Washizaki, H. and Maruyama, K.,** 2008. A survey on security patterns. Progress in informatics, 5(5), pp.35-47.

[34] **Cardenas, A.A., Roosta, T. and Sastry, S.,** 2009. Rethinking security properties, threat models, and the design space in sensor networks: A case study in SCADA systems. Ad Hoc Networks, 7(8), pp.1434-1447.

[35] **Blair, M.M., Williams, C.A. and Lin, L.W.,** 2007. The new role for assurance services in global commerce. J. Corp. L., 33, p.325.

[36] **Taylor, D.E.,** 2005. Survey and taxonomy of packet classification techniques. ACM Computing Surveys (CSUR), 37(3), pp.238-275.

[37] **Vimina, E.R. and Areekal, A.U.,** 2009, October. Telling computers and humans apart automatically using activity recognition. In Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on (pp. 4906-4909). IEEE.

[38] **Stiawan, D., Abdullah, A.H. and Idris, M.Y.,** 2010, June. The trends of intrusion prevention system network. In Education Technology and Computer (ICETC), 2010 2nd International Conference on (Vol. 4, pp. V4-217). IEEE.

[39] **Zhang, Y., Deng, F., Chen, Z., Xue, Y. and Lin, C.,** 2010, April. UTM-CM: A practical control mechanism solution for UTM system. In Communications and Mobile Computing (CMC), 2010 International Conference on (Vol. 1, pp. 86-90). IEEE.

[40] **Conklin, W.A. and Dietrich, G.,** 2008, January. Systems theory model for information security. In Hawaii International Conference on System Sciences, Proceedings of the 41st Annual (pp. 265-265). IEEE.

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

[42] **K. Balan, R. Ramasamy** 2015. "A Text and Image Recognition CAPTCHA-A Cognitive Authentication Scheme Based On Hard AI Problems"

[43] **Daş, R., Baykara, M., Tuna, G.,** 2015. "Novel CAPTCHA Approaches to Protect Web Forms against Bots", The Third International Symposium on Digital Forensics and Security (ISDFS 2015), 11-12 May, 2015, Gazi University, ANKARA.

[44] **Yeşiltepe, M.,** 2016. Appropriate User Acceptance Criteria For New Social Media

[45] **Mori, G. and Malik, J.,** 2003, June. Recognizing objects in adversarial clutter: Breaking a visual CAPTCHA. In Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on (Vol. 1, pp. I-I). IEEE.

[46] **Gao, H., Wang, W., Fan, Y., Qi, J., & Liu, X.** 2014. The Robustness of" Connecting Characters Together" CAPTCHAs. J. Inf. Sci. Eng., 30(2), 347-369.

[47] **Kluever, K.A.,** 2008. Evaluating the Usability and Security of a Video CAPTCHA.

[48] **Zhang, W.,** 2010, April. Zhang's CAPTCHA architecture based on intelligent interaction via RIA. In Computer Engineering and Technology (ICCET), 2010 2nd International Conference on (Vol. 6, pp. V6-57). IEEE.

[49] **Saini, B.S. and Bala, A.,** 2013. A review of bot protection using CAPTCHA for web security. IOSR Journal of Computer Engineering, 6, pp.36-42..

[50] **Rao, M.K., Maniraj, M.S.V.K. and Ganga, B.S.,** 2014. Improved Video Captcha. Journal of Emerging Technologies in Web Intelligence, 6(4), pp.416-420.

[51] **Shirali-Shahreza, M. and Shirali-Shahreza, S.,** 2007, December. Question-based CAPTCHA. In Conference on Computational Intelligence and Multimedia Applications, 2007. International Conference on (Vol. 4, pp. 54-58). IEEE.

[52] **Chen, C. J., Wei, Y., & Fang, W. P.** 2014, August. A Study on Captcha Recognition. In Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference on (pp. 395-398). IEEE.

[53] **Ng, P.E. and Ma, K.K.,** 2006. A switching median filter with boundary discriminative noise detection for extremely corrupted images. IEEE Transactions on image processing, 15(6), pp.1506-1516.

[54] **Gonzalez, R.C.,** 2008. Digital image processing, RC Gonzalez and RE Woods, Eds.

[55] **Baecher, P., Büscher, N., Fischlin, M. and Milde, B.,** 2011. Breaking reCAPTCHA: a holistic approach via shape recognition. Future challenges in security and privacy for academia and industry, pp.56-67.

[56] **Chellapilla, K., Larson, K., Simard, P.Y. and Czerwinski, M.,** 2005. Building Segmentation Based Human-Friendly Human Interaction Proofs (HIPs). HIP, 3517, pp.1-26.

[57] **Greg, M, Jitendra, M.** 2003. Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 134-141.

[58] **Ghorpade, J., Mukane, S., Patil, D., Poal, D. and Prasad, R.,** 2014. Novel Method for Graphical Passwords using CAPTCHA. International Journal of Soft Computing and Engineering (IJSCE) ISSN, pp.2231-2307.

[59] **Alginahi, Y.,** 2010. Preprocessing techniques in character recognition (pp. 1-20). INTECH Open Access Publisher.

[60] **Sezgin, M.,** 2004. Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic imaging, 13(1), pp.146-168.

[61] **Gupta, M.R., Jacobson, N.P. and Garcia, E.K.,** 2007. OCR binarization and image pre-processing for searching historical documents. Pattern Recognition, 40(2), pp.389-397.

[62] **Yan, J. and El Ahmad, A.S.,** 2007. December. Breaking visual captchas with naive pattern recognition algorithms. In Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual (pp. 279-291). IEEE.

[63] **Eng, H.L. and Ma, K.K.,** 2001. Noise adaptive soft-switching median filter. IEEE Transactions on image processing, 10(2), pp.242-251.

CURRICULUM VITA

Karmand Hussein Abdalla

E-mail karmand.hussen@yahoo.com

Nationality: Iraqi

Place of birth: Sulaimanyah-Iraq

Date of birth: 22 / 05 / 1988

Marital status: Married

EDUCATION

09/ 2015 – Present MSc (Master of Software Engineering), Firat University, Elazig, Turkey.

09/ 2007 – 07/ 2011 Bachelor degree from Soran University, College of Science, Computer Science department. Iraq, Erbil-Soran.

09/ 2006 – 07/2007 higher School from Qaladza Preparatory for boys, Iraq, Sulaimanyah

WORK EXPERIENCES

1. Assistant Programmer at Soran University. 11/2011-07/2014
2. Teacher Computer Science at Soran Nongovernmental Computer Institute. 10/2013 06/2014
3. IT Support at Tishknet Company in Sulaimanyah-Ranya 03/2015-09/2015
4. Assistant Programmer at University of Raparin 07/2014-present

PUBLICATIONS

Abdalla, K. and Kaya, M., 2017. AN EVALUATION OF DIFFERENT TYPES OF CAPTCHA: EFFECTIVENESS, USER-FRIENDLINESS, AND LIMITATIONS. International Journal of Scientific Research in Information Systems and Engineering (IJSRISE), 2(3).