

**ISAS**

**“reCAPTCHA on Google"**

Arranged by:

1. Kapten Bintang
2. Muhamad Dani Setiawan

Faculty:

Riza Muhammad Nurman S.Kom

**CONTINUING EDUCATION PROGRAM CENTER FOR COMPUTING AND INFORMATION TECHNOLOGY**

**FACULTY OF ENGINEERING, UNIVERSITY OF INDONESIA**

**2020**

**PREFACE**

Praise God Almighty, for the presence of plenty of mercy and his grace, so that we can complete this Information Search and Analysis Skills (ISAS) with the title “reCAPTCHA on Google”. Without His mercy, we would not be able to complete this project in time. Even though there are many obstacles that author face on making this project, but finally author can finish this project.

The author also thanked Mr. Riza Muhammad Nurman as a lecturer who has provided guidance to the author and advice in the process of preparing this ISAS. Not to forget the author thanked the various parties who have provided encouragement and motivation so that the ISAS can be completed on time

Author know that the results of this article is far from perfect and there are still many shortcomings, author hope readers will give comments and suggestions in building this article in order to become better. We hope this article can be useful for those who read or hear, especially for CCIT students of the Faculty of Engineering UI.

Thank you,

Depok,9 June 2020

Author

**TABLE OF CONTENTS**

**PREFACE i**

**TABLE OF CONTENTS ii**

**TABLE OF FIGURE iii**

**CHAPTER I INTRODUCTION 1**

1. Background 1
2. Writing Objective 2
3. Problem Domain 2
4. Writing Methodology 3
5. Writing Framework 3

**CHAPTER II BASIC THEORY 4**

1. Explanation of Cyber Security 4
2. Explanation of Bot Attacks 4
3. What is CAPTCHA 6
4. What is reCAPTCHA 6
5. Types of reCAPTCHA 7

**CHAPTER III PROBLEM ANALYSIS 8**

1. How Does reCAPTCHA Works 9
2. Advantages and Disadvantages of using reCAPTCHA 10
3. An Architecture for reCAPTCHA on Google 11

**CHAPTER IV CONCLUSION & SUGGESTION 13**

4.1 Conclusion 13  
4.2 Suggestion 13

**BIBILIOGRAPHY iv**

**TABLE OF FIGURES**

Figure 3.1 How Does reCAPTCHA works 8

Figure 3.2 Checkbox reCAPTCHA 9

Figure 3.3 Image Verification reCAPTCHA 10

Figure 3.4 Architecture reCAPTCHA 11

**CHAPTER I  
INTRODUCTION**

1. **Background**

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is a test used in computing to distinguish whether or not the user is computer or human. Many web sites mostly use the character-based CAPTCHA consisting of digits and characters.

The original iteration of the service was a mass collaboration platform designed for the digitization of books, particularly those that were too illegible to be scanned by computers. The verification prompts utilized pairs of words from scanned pages, with one known word used as a control for verification, and the second used to crowdsource the reading of an uncertain world. reCAPTCHA was originally developed by Luis von Ahn, David Abraham, Manuel Blum, Michael Crawford, Ben Maurer, Colin McMillen, and Edison Tan at Carnegie Mellon University's main Pittsburgh campus. It was acquired by Google in September 2009.

The CAPTCHA was created at Carnegie Mellon University in 2000. The name is short for Completely Automated Public Turing test to tell Computers and Humans Apart. Websites need CAPTCHAs to guard against the "bots" of spammers and other computer underworld types. "Anybody can write a program to sign up for millions of accounts, and the idea was to prevent that," said Luis von Ahn, a Carnegie Mellon professor who was part of the CAPTCHA team. The little puzzles work because computers are not as good as humans at reading distorted text. Google says that people are solving 200 million CAPTCHA a day.

In 2014, a Google analysis found that artificial intelligence could crack even the most complex CAPTCHA and reCAPTCHA images with 99.8 percent accuracy, rendering the programs useless as security devices. In their place, Google unveiled the now-familiar “No CAPTCHA reCAPTCHA” system, which relies not on a users ability to decipher the text, but on their online behavior prior to the security checkpoint.

1. **Writing Objective**
2. Explanation of Cyber Security
3. Explanation of Bot Attacks
4. What is CAPTCHA
5. What is reCaptcha
6. Types of reCaptcha
7. **Problem Domain**

Accordance with the title of ISAS "reCAPTCHA on Google" We will discuss:

1. How Does reCAPTCHA Works
2. Advantages and Disadvantages of using reCAPTCHA
3. An Architecture for reCAPTCHA
4. **Writing Methodology**

**Library Research Method**

Collecting data with browsing information from reference source contain on online sites that relate with the topic of this ISAS.

**Discussion Method**

After collecting data from reference source, we discuss and compose the data into structure contents for completing this ISAS.

1. **Writing Framework**
   * 1. **Chapter I Introduction**

* Background
* Writing Objectives
* Problem Domain
* Methodology
* Writing Framework
  + 1. **Chapter II Basic Theory**
* Explanation of Cyber Security
* Explanation of Bot Attacks
* What is CAPTCHA
* What is reCAPTCHA
* Types of reCAPTCHA
  + 1. **Chapter III Problem Analysis**
* How Does reCAPTCHA Works
* Advantages of using reCAPTCHA on Google
* An Architecture for reCAPTCHA
  + 1. **Chapter IV Conclusion**
* Conclusion
* Suggestion

**CHAPTER II**

**BASIC THEORY**

1. **Explanation of Cyber Security**

Cyber Security Security is the protection of internet-connected systems, including hardware, software and data, from cyberattacks. In a computing context, security comprises cybersecurity and physical security -- both are used by enterprises to protect against unauthorized access to data centers and other computerized systems. The goal of cybersecurity is to limit risk and protect IT assets from attackers with malicious intent. Information security, which is designed to maintain the confidentiality, integrity and availability of data, is a subset of cybersecurity.

Cybersecurity best practices can, and should, be implemented by large and small organizations, employees and individuals. One of the most problematic elements of cybersecurity is the continually evolving nature of security risks and advanced persistent threats (APTs).

The traditional approach has been to focus resources on crucial system components and protect against the biggest known threats, which meant leaving components undefended and not protecting systems against less dangerous risks. To deal with the current environment, advisory organizations are promoting a more proactive and adaptive approach.

The National Institute of Standards and Technology (NIST), for example, recently issued updated guidelines in its risk assessment framework that recommend a shift toward continuous monitoring and real-time assessments. Version 1.1 of the Framework for Improving Critical Infrastructure was released in April 2018. The voluntary Cybersecurity Framework (CSF), developed for use in the banking, communications, defense and energy industries, can be adopted by all sectors, including federal and state governments. President Donald Trump issued an executive order mandating that federal agencies adopt the NIST CSF in May 2017.

1. **Explanation of Bot Attacks**

A bot is a software application that is programmed to do certain tasks. Bots are automated, which means they run according to their instructions without a human user needing to start them up.Typically they do repetitive tasks, and they can do them much faster than human users could.

Bots usually operate over a network; more than half of Internet traffic is scanning content, interacting with webpages, chatting with users, or looking for attack targets. Some bots are useful, such as search engines bots that index content for search or customer service bots that help users. Other bots are "bad" and are programmed to break into user accounts, scan the web for contact information for sending spam, or perform other malicious activities. If it's connected to the Internet, a bot will have an associated IP address.

What is a malicious bot activity?

Any automated actions by a bot that violate a website owner's intentions, the site's Terms of Service, or the site's Robots.txt rules for bot behavior can be considered malicious. Bots that attempt to carry out cybercrime, such as identity theft or account takeover, are also "bad" bots. While some of these activities are illegal, bots do not have to break any laws to be considered malicious.

In addition, excessive bot traffic can overwhelm a web server's resources, slowing or stopping service for the legitimate human users trying to use a website or an application. Sometimes this is intentional and takes the form of a DoS or DDoS attack

**2.3 What is CAPTCHA**

A CAPTCHA test is designed to determine if an online user is really a human and not a bot. CAPTCHA is an acronym that stands for "Completely Automated Public Turing test to tell Computers and Humans Apart." Users often encounter CAPTCHA and reCAPTCHA tests on the Internet. Such tests are one way of managing bot activity, although the approach has its drawbacks.

Although CAPTCHAs are designed to block automated bots, CAPTCHAs are themselves automated. They're programmed to pop up in certain places on a website, and they automatically pass or fail users. The idea is that a computer program such as a bot will be unable to interpret the distorted letters, while a human being, who is used to seeing and interpreting letters in all kinds of contexts different fonts, different handwritings, etc.

CAPTCHAs prevent spammers from using bots to create a plethora of spam email accounts. Web pages or blogs that contain message boards or contact forms use CAPTCHA to prevent spammy messages or comments. It does not prevent cyberbullying but does prevent bots from posting messages automatically. The best that many bots will be able to do is input some random letters, making it statistically unlikely that they will pass the test. This, bots fail the test and are blocked from interacting with the website or application, while humans are able to continue using it like normal.

Advanced bots are able to use machine learning to identify these distorted letters, so these kinds of CAPTCHA tests are being replaced with more complex tests. Google reCAPTCHA has developed a number of other tests to sort out human users from bot.

**2.4 What is reCAPTCHA**

reCAPTCHA is a free service Google offers as a replacement for traditional CAPTCHAs. reCAPTCHA technology was developed by researchers at Carnegie Mellon University, then acquired by Google in 2009.

reCAPTCHA is more advanced than the typical CAPTCHA tests. Like CAPTCHA, some reCAPTCHA require users to enter images of text that computers have trouble deciphering. Unlike regular CAPTCHAs, reCAPTCHA sources the text from real-world images: pictures of street addresses, text from printed books, text from old newspapers, and so on.

Over time, Google has expanded the functionality of reCAPTCHA tests so that they no longer have to rely on the old style of identifying blurry or distorted text. Other types of reCAPTCHA tests include:

* Image recognition
* Checkbox
* General user behavior assessment

**2.5 Types of reCAPTCHA**

Until now reCAPTCHA has three versions, namely, reCAPTCHA v1, v2, and v3. reCAPTCHA v1 is a form to complete various letter codes (CAPTCHA). While v2 looks with a newer look. In general, reCAPTCHA v2 has the words "I am not a robot", and the user just need to check the list, reCAPTCHA uses IE / Edge / Chrome on Windows and Safari / Chrome on Mac OS.

reCAPTCHA v3 allows you to verify if an interaction is legitimate without any user interaction. It is a pure JavaScript API returning a score, giving you the ability to take action in the context of your site: for instance requiring additional factors of authentication, sending a post to moderation, or throttling bots that may be scraping content.

reCAPTCHA v2, "I'm not a robot" Checkbox requires the user to click a checkbox indicating the user is not a robot. This will either pass the user immediately (with No CAPTCHA) or challenge them to validate whether or not they are human. This is the simplest option to integrate with and only requires two lines of HTML to render the checkbox. reCAPTCHA v1 has been closed since March 2018.

**CHAPTER III**

**PROBLEM ANALYSIS**

**3.1 How Does reCAPTCHA Works**



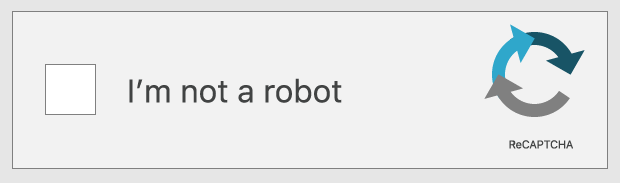
**Figure 3.1 How does reCAPTCHA Works   
(https://www.kagetweb.com/image/data/blog/recaptcha\_hard\_april\_2.jpg)**

How does reCAPTCHA work? The book digitization project was converted to text using Optical Character Recognition (OCR). The process of change is needed to convert a book that has an image format to text so that its size will be lighter and also easier to read. OCR itself is software that is not perfect because it is not able to solve all the characters correctly. it is reCAPTCHA's job to be able to help it. reCAPTCHA helps the process of digitizing books by sending words that cannot be recognized by a computer, to a website in the form of a CAPTCHA. This program is possible because almost all OCR programs can distinguish whether words can be read well or not.

How does the computer know that the keywords we input are real words? Any words that cannot be read correctly by the OCR program will be combined with words that have been recognized as true. The user is then asked to answer it.If they can solve words that have proven the truth, the system will answer the correct answer to other words. That is the secret   
of reCAPTCHA which consists of two words.

**3.1.1 How does reCAPTCHA tests with a single checkbox work**

Some reCAPTCHA tests simply prompt the user to check a box next to the statement, "I'm not a robot." However, the test is not the actual action of clicking the checkbox it's everything leading up to the checkbox click.



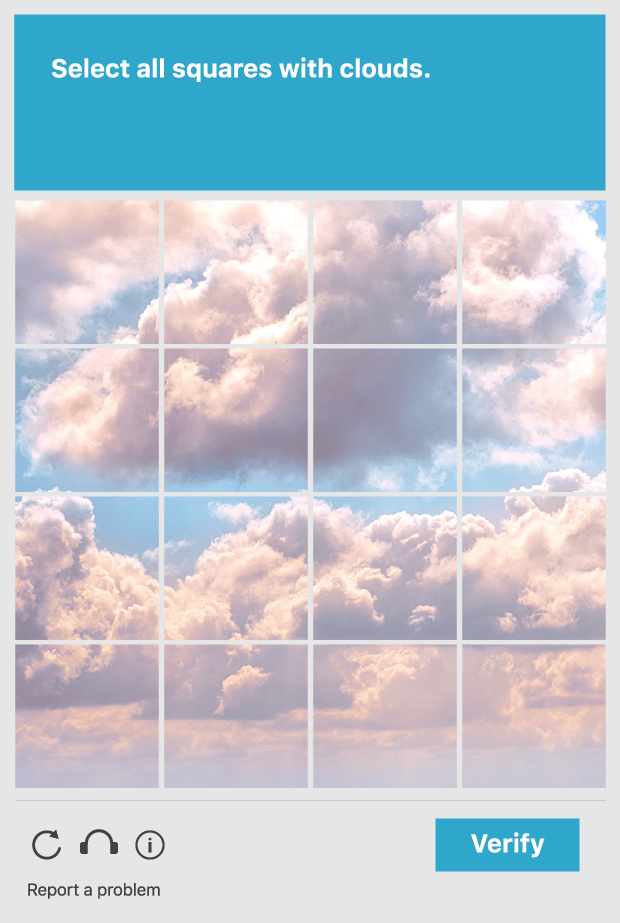
**Figure 3.2 Checkbox reCAPTCHA  
(https://www.cloudflare.com/resources/images/slt3lc6tev37/79wsjD0Xy7FmmYvR0sCncy/5b732b7e26adb7d6c06d943d14dc4acd/not-a-robot.png)**

This reCAPTCHA test takes into account the movement of the user's cursor as it approaches the checkbox. Even the most direct motion by a human has some amount of randomness on the microscopic level: tiny unconscious movements that bots can't easily mimic. If the cursor's movement contains some of this unpredictability, then the test decides that the user is probably legitimate. The reCAPTCHA also may assess the cookies stored by the browser on a user device and the device's history in order to tell if the user is likely to be a bot.

If the test is still unable to determine whether or not the user is a human, it may present an additional challenge. However, most of the time the user's cursor movements, cookies, and device history are conclusive enough.

**3.1.2 How does an image recognition reCAPTCHA test work?**

For an image recognition reCAPTCHA test, typically users are presented with 9 or 16 square images. The images may all be from the same large image, or they may each be different. A user has to identify the images that contain certain objects, such as animals, trees, or street signs. If their response matches the responses from most other users who have submitted the same test, the answer is considered "correct" and the user passes the test



**Figure 3.3 reCAPTCHA Image Verification  
(https://www.cloudflare.com/resources/images/slt3lc6tev37/3mWhV5AJh3QeW7ZCsEHm9N/1104811b27166f82bce8157a5b974183/image-captcha.png)**

Picking out certain objects from blurry photos is a hard problem for computers to solve. Even advanced artificial intelligence (AI) programs struggle with it, so a bot will struggle with it as well. However, a human user should be able to do this fairly easily, since humans are used to perceiving everyday objects in all kinds of contexts and situations.

**3.2 Advantages and Disadvantages of using reCAPTCHA**

Captcha is a way to differentiate between an automated computer program and a human. It is a box with distorted text that must be deciphered in order to enter free email services, online polls, and to complete online purchases.

Advantages:

* Distinguishes between a human and a machine
* Makes online polls more legitimate
* Reduces spam and viruses
* Makes online shopping safer
* Diminishes abuse of free email account services

Disadvantages:

* Sometimes very difficult to read
* Are not compatible with users with disablilities
* Time-consuming to decipher
* Technical difficulties with certain internet browsers
* May greatly enhance Artificial Intelligence

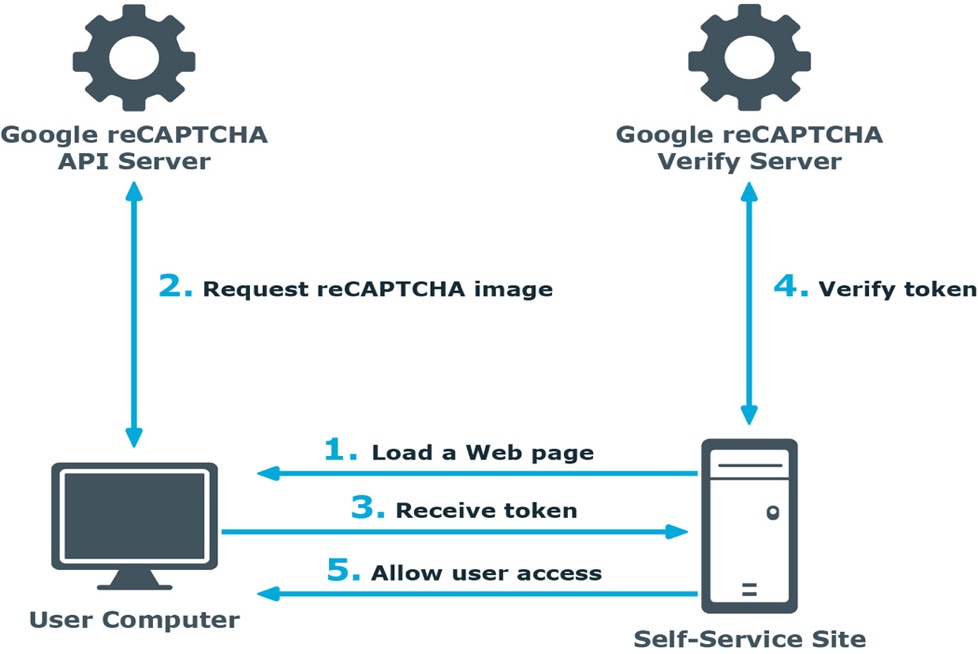
**3.2.1 What are the drawbacks of using reCAPTCHA to stop bots**

**-**Bad user experience: A CAPTCHA test can interrupt the flow of what users are trying to do, giving them a negative view of their experience on the web property, and leading to them abandoning the webpage altogether in some cases.

-Not usable for visually impaired individuals: The problem with CAPTCHAs is that they rely on visual perception. This makes them nearly impossible, not just for people who are legally blind, but for anyone with seriously impaired vision.

-These tests can be fooled by bots: CAPTCHAs are not fully bot-proof and shouldn't be relied upon for bot management.

**3.3 An Architecture for reCAPTCHA**



**Figure 3.4 Architecture reCAPTCHA**

**(**[**https://www.researchgate.net/figure/Web-Application-Firewall-architecture\_fig1\_267414465**](https://www.researchgate.net/figure/Web-Application-Firewall-architecture_fig1_267414465)**)**

1. A user opens the Self-Service site.
2. The user’s browser sends the site key obtained during registration on the reCAPTCHA V2 site to the Google reCAPTCHA V2 API server and requires the user to click a checkbox indicating the user is not a robot.
3. Use this activity to verify reCAPTCHA on the Self-Service site. User must click on the I'm not a robot check box before beginning a workflow. This will either pass the user immediately (with No CAPTCHA) or challenge them to validate whether or not they are human. This feature provides enhanced protection against automated attacks.
4. The token and the secret key (obtained during registration on the reCAPTCHA V2 site) are then transferred to the Google reCAPTCHA V2 Verify server to be checked. After checking the response, the reCAPTCHA V2 server sends a reply back to the Password Manager server.
5. If the response is correct, the user is granted access to further steps on the Password Manager site.

**CHAPTER IV**

**CONCLUSION AND SUGGESTION**

1. **Conclusion**

Unfortunately, as bots get more and more advanced, reCAPTCHA methods will have to become more and more intrusive to establish that you are in fact human. Users will need to sacrifice more of their privacy in order to remain safe from spam and bots on the internet, and by the looks of things, it’s something we’ve got to accept.

1. **Suggestion**

We recommend to be more careful in computer network systems, and we recommend that you always use a security protection system to prevent attacks on networks such as this and always follow the development of any security system in the world to always be safe.

# Bibliography

**1.Shah,Bhargav.**2012. https://medium.com/@bhargavshah2011/basics-of-recaptcha-v3-the-new-way-to-stop-bots-5e14054c4026 [08/06/2020]

**2. Overson,Jarrod**.2018. https://medium.com/@jsoverson/thoughts-on-recaptcha-v3-e837d4a0a63 [08/06/2020]

**3. Ego,Alter**.2017. https://medium.com/@cyberalterego/everything-you-need-to-know-about-captcha-f77d65c56d2c [08/06/2020]

**4.** https://www.cloudflare.com/learning/bots/how-captchas-work/ [12/06/2020]

**5.** https://www.cloudflare.com/learning/bots/what-is-spambot/ [12/06/2020]

**6.Owen,Williams.**2019. https://onezero.medium.com/google-promises-recaptcha-isn-t-exploiting-users-should-you-trust-it-ed99f1543f28 [14/06/2020]