# University of Padova
## Engineering Departement

*Computer Engineering Master Degree*

# Invisible CAPPCHA

*Grad Student*

**Di Nardo Di Maio Raffaele**

*Supervisor*

**Prof. Migliardi Mauro**

*Co-Supervisors*

**Guerar Meriem**

DD-MM-YYYY

Accademic Year 2020-2021

ii

To my parents, that always help
me to be happy doing what I love
and support me reaching my goals.

# Contents

# Chapter 1

# Introduction

CAPTCHA (Completely Automated Public Turing Test to Tell Computers and Humans Apart) is a program used to distinguish human users from bots. A bot is a malicious application that automates a task, gathering useful information about user credentials or pretending to be a human interaction with Web application.

The CAPTCHAs are traditionally used in Web applications for[3]:

- **Online Polls**
  CAPTCHAs prevent the creation and the submission of a large number of votes, favouring a party.

- **Protecting Web Registration**
  CAPTCHAs prevent the creation of free mail account to bot instead of human users. The goal of the use of CAPTCHAs is to remove the possibility that the hacker could take advantages from the large amount of registrations.

- **Preventing comment spam**
  CAPTCHAs prevent the insertion of a large amount of posts made by bot on pages of social platforms or blogs.

- **Search engine bots**
  CAPTCHAs are used to guarantee that a website should be unindexed to prevent the reading of the page through search engine bots. The CAPTCHAs are added because the html tag, used to unindex the web page, doesn't guarantee unindexing.

- **E-Ticketing**
  CAPTCHAs prevent that a big events would sell out minutes after

tickets become available. In fact ticket scalpers that make large number
of ticket purchases for big events.

- **Email spam**
  CAPTCHAs are used to verify that a human has sent the email.

- **Preventing Dictionary Attacks**
  CAPTCHAs prevent bot to guess the password of a specific user. The
  hacker could guess the password, taking it from a dictionary of pass-
  words. The use of the CAPTCHA challenge prevents the iteration of
  the login phase made by the bot using all the words of the dictionary.
  After a certain number of failures POST requests, the CAPTCHA chal-
  lenge is shown to the user.

- **Verifying digitized books**
  **DA RIVEDERE, NON HO CAPITO**
  This is a way of increasing the value of CAPTCHA as an applica-
  tion. An application called reCAPTCHA harnesses users responses in
  CAPTCHA fields to verify the contents of a scanned piece of paper.
  Because computers aren't always able to identify words from a digital
  scan, humans have to verify what a printed page says. Then it's pos-
  sible for search engines to search and index the contents of a scanned
  document. This is how it works: The application already recognizes
  one of the words. If the visitor types that word into a field correctly,
  the application assumes the second word the user types is also correct.
  That second word goes into a pool of words that the application will
  present to other users. As each user types in a word, the application
  compares the word to the original answer. Eventually, the applica-
  tion receives enough responses to verify the word with a high degree of
  certainty. That word can then go into the verified pool.

Another useful application of CAPTCHA is the support to the authentica-
tion process. This application is going to be analysed in details in the next
chapters, looking at the authentication from smartphone. **Description of
the chapters CONTENT**

# Chapter 2

# State of Art

## 2.1 History

CAPTCHA is related to three main elements[4]:

1. **Turing test**

2. **Human-Computer Interaction (HCI)**

3. **Human Interactive Proof (HIP)**

## 2.2 Traditional CAPTCHA

The following types of CAPTCHA are the challenges most related to the historical Turing test:

- **Arithmetic**

- **Audio-based**

- **Game-based**

- **Image-based**

- **Puzzle-based**

- **Text-based**

- **Video-based**

Some types of CAPTCHA don't destroy a session, after the correct answer is inserted by the user[3]. Hence, the hacker can crack following accesses using the same session id with the related solution of the challenge, after connecting to the web page of CAPTCHA. In this way the attacker can make hundreds of requests before the session expires and the previous operation must be computed again.

## 2.3   Alternatives

This types of CAPTCHA and authentication mechanisms are far from traditional CAPTCHAs and aren't based on cognitive knowledge of the human user but on other parameters:

- **Biometrics-based**

- **Behavioural-based**

- **Social media sign-in**

| CAPTCHA type | Usability issues | Security |
|---|---|---|
| *Arithmetic* | | |
| *Audio-based* | Issues of recognition:<br>• Previous knowledge of English dictionary by the user.<br>• Some character sounds very similar to others. | It can be broken by Automatic Speech Recognition (ASR) programs (as mentioned in [2]). |
| *Game-based* | | |
| *Image-based* | Difficulty of identification of images caused by:<br>• Blur of images.<br>• Low vision condition. | |
| *Puzzle-based* | It takes too much time to solve the puzzle and to identify the arrangement of puzzles. | |
| *Text-based* | Many problems have to be solved by user:<br>• Multiple fonts.<br>• Font size.<br>• Blurred Letters<br>• Wave Motion. | It can be identified by:<br>• OCR (Optical Character Recognition) technique<br>• Segmentation techniques (e.g. DECAPTCHA[1])<br>• Machine Learning and Deep Learning techniques |
| *Video-based* | Issues downloading videos to find correct captcha because of large size of files. | |

# Bibliography

[1] E. Bursztein, M. Martin, and J. Mitchell, "Text-based CAPTCHA strengths and weaknesses" in *Proc. 18th ACM Conf. Comput. Commun. Secur. (CCS), 2011, pp. 125–138.*

[2] Jennifer Tam, Jiri Simsa, David Huggins-Daines, Luis von Ahn, and Manuel Blum, "Improving Audio CAPTCHAs" in *Symposium On Usable Privacy and Security (SOUPS), 2008.*

[3] Sarika Choudhary, Ritika Saroha, Yatan Dahiya, and Sachin Choudhary, "Understanding CAPTCHA: Text and Audio Based Captcha with its Applications" in *International Journal of Advanced Research in Computer Science and Software Engineering vol. 3 (6), pp. 106-115, June-2013.*

[4] Brodi'c, D.; Amelio, A.; Jankovi'c, R.; "Exploring the influence of CAPTCHA types to the users response time by statistical analysis" in *Multimed. Tools Appl. 2017, 77, 12293–12329*