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

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

Contents

Chapter 1

Introduction

The CAPTCHAs are traditionally used in Web applications for:

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

2. **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.

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

4. **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.

5. **E-Ticketing**
   Ticket brokers like Ticketmaster also use CAPTCHA applications. These applications help prevent ticket scalpers from bombarding the service with massive ticket purchases for big events. Without some sort of filter, it's possible for a scalper to use a boot to place hundreds or thousands of ticket orders in a matter of seconds. Legitimate customers become victims as events sell out minutes after tickets become available. Scalpers then try to sell the tickets above face value. While

CAPTCHA applications don't prevent scalping; they do make it more difficult to scalp tickets on a l

6. **Email spam**
   CAPTCHAs also present a plausible solution to the problem of spam emails. All we have to do is to use a CAPTCHA challenge to verify that an indeed a human has sent the email.

7. **Preventing Dictionary Attacks**
   CAPTCHAs can also be used to prevent dictionary attacks in password systems. The idea is simple: prevent a computer from being able to iterate through the entire space of passwords by requiring it to solve a CAPTCHA after a certain number of unsuccessful logins. This is better than the classic approach of locking an account after a sequence of unsuccessful logins, since doing so allows an attacker to lock accounts at will.

8. **As a tool to verify digitized books: This is a way of increasing the value of CAPTCHA as an application. 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 possible 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 application receives enough responses to verify the word with a high degree of certainty. That word can then go into the verified pool.**

1.1 Traditional CAPTCHA

- **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.

1.2 Alternatives

- **Biometrics-based**
- **Behavioural-based**
- **Social media sign-in**

CAPTCHA type	**Usability issues**	**Security**
Arithmetic		
Audio-based	Issues of recognition: • Previous knowledge of English dictionary by the user. • Some character sounds very similar to others.	It can be broken by Automatic Speech Recognition (ASR) programs (as mentioned in [?]).
Game-based		
Image-based	Difficulty of identification of images caused by: • Blur of images. • 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: • Multiple fonts. • Font size. • Blurred Letters • Wave Motion.	It can be identified by: • OCR (Optical Character Recognition) technique • Segmentation techniques (e.g. DECAPTCHA[1]) • 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.*