



UNIVERSITY OF PADOVA
ENGINEERING DEPARTEMENT

Computer Engineering Master Degree

ACCAPPCHA:
DESIGN, DEVELOPMENT AND SECURITY ANALYSIS
OF AN INVISIBLE CAPTCHA BASED ON
AN ACOUSTIC SIDE-CHANNEL

Candidate

Di Nardo Di Maio Raffaele

Supervisor

Prof. Migliardi Mauro

DD-MM-2021

ACADEMIC YEAR 2021-2022

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

“Most people assume that once security software is installed, they’re protected. This isn’t the case. It’s critical that companies be proactive in thinking about security on a long-term basis.”

Kevin Mitnick

“You have to learn the rules of the game. And then you have to play better than anyone else.”

Albert Einstein

“Si come il ferro s’arrugginisce senza esercizio, e l’acqua si putrefà o nel freddo s’addiaccia, così lo ’ngegno senza esercizio si guasta.”

Leonardo da Vinci

Abstract

Sommario

Contents

Abstract	vii
1 Introduction	1
2 State of the Art	5
2.1 Traditional CAPTCHAs	7
2.1.1 Audio-based CAPTCHAs	7
2.1.2 Game-based CAPTCHAs	8
2.1.3 Image-based CAPTCHAs	8
2.1.4 Math CAPTCHAs	12
2.1.5 Slider CAPTCHAs	12
2.1.6 Text-based CAPTCHAs	13
2.1.7 Video-based CAPTCHAs	14
2.2 Modern CAPTCHAs	15
2.2.1 Biometrics-based CAPTCHAs	15
2.2.2 Behavioural-based CAPTCHAs	16
2.2.3 Sensor-based CAPTCHAs	18
2.3 CAPTCHA security	19
3 Side-channel attacks	23
3.1 Local side-channel attacks	24
3.2 Vicinity side-channel attacks	26
3.3 Remote side-channel attacks	27
4 Invisible CAPPCHA	31
4.1 Motion detection	31
4.2 Communication between Client and Server	33
4.2.1 Elliptic Curve Digital Signature Algorithm (ECDSA) .	33
4.3 Security analysis	35
4.3.1 Strength against popular attacks	35

5 AcCAPPCHA	37
5.1 Evaluation of the user's activity	38
5.1.1 Time correspondence	39
5.1.2 Character correspondence	42
5.1.2.1 Data acquisition	43
5.1.2.2 Classification	45
5.1.2.3 Verification	48
5.2 Communication between client and server	48
5.2.1 Client	49
5.2.2 Server	51
5.2.3 Database	53
5.2.4 Encryption Keys	55
6 Experimental results	57
6.1 Human detection	57
6.2 Bot detection	59
6.3 Usability	61
6.4 Security analysis	63
7 Future work	65
A Key Map	67
B Program	75
Bibliography	77

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 do a human interaction with a Web application. In fact the term "*bot*" is an abbreviation of the words "software robot".

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.

- **Protecting Web Registration**

CAPTCHAs prevent the creation of free mail account for bots instead of human users. The aim of CAPTCHAs is to remove the possibility that the hacker could take advantages from a large amount of registrations.

- **Preventing comment spam**

CAPTCHAs prevent the insertion of many posts made by a bot on pages of social platforms or blogs.

- **Search engine bots**

CAPTCHAs are used to guarantee that a website would be unindexed and to prevent search engine bots from reading a page. 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 event would sell out minutes after the tickets become available. In fact CAPTCHAs reduces the number of

scalpers that make many ticket purchases and sell them with higher prices.

- **Email spam**

CAPTCHAs are used to verify that a human has sent an email.

- **Preventing Dictionary Attacks**

CAPTCHAs prevent bot from guessing the password of a specific user. The hacker could guess the password, taking it from a dictionary. 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 challenge is shown to the user.

- **Verifying digitized books**

ReCAPTCHA can verify the contents of a scanned piece of paper analysing responses in CAPTCHA fields. A computer can't identify all the words from a digital scan.

The application submits two words to the user during the CAPTCHA challenge: the first one that the machine has already recognized and the other one for which the application wants to associate a word. If the user types the two words and the first one was correctly detected, it assumes that also the second one is correct.

In this case the second word is added to a set of words that are going to be choose to create challenges for other users. If the application receives enough responses with the same typed word related to an unknown word, the program establishes that that label would be associated to it. Hence the CAPTCHA scans digitally the paper during the verification of the user identity.

However, over the years, these challenges become more and more complex because a lot of them easily fail against new machine learning techniques (**Chapter 2**).

Many attacks exploits data, collected by built-in sensors of the victim's machine, to obtain some credentials or useful information (**Chapter 3**). Hence, in the last years, the sensors were exploited not only for malicious purposes but also to increase the strength of CAPTCHAs against bot attacks.

CAPTCHAs are also very used during the authentication process. For example, Invisible CAPTCHA is a new CAPTCHA scheme, developed in 2018, that exploits the physical information obtained by motion sensors of the smartphone. Using this CAPTCHA, the user simply authenticates him self and the application analyses in background if this phase was performed by a

human (**Chapter 4**).

From the ideas developed in Invisible CAPPCHA and exploiting the acoustic side-channel, I designed AcCAPPCHA (**Chapter 5**) to support the authentication from the desktop and the laptop environments.

All the pros and the cons of the new invisible challenge, observed during the testing phase, have been described (**Chapter 5**) and all the ideas of future improvements are analysed in (**Chapter 7**).

In the appendices of the current work, you can also find the key map used by AcCAPPCHA (**Appendix A**) and the details about its command line arguments (**Appendix B**).

Chapter 2

State of the Art

CAPTCHA is the acronym of "*Completely Automated Public Turing-test-to-tell Computers and Humans Apart*" and it's a challenge born from the union of the following tests[2]:

1. Turing test

it's a test used to determine how much a machine could think like a human. The experiment involves three figures: a human examiner, a human and a machine. The examiner asks some questions to the other figures and, after a fixed amount of time, he evaluates if the two answers were different or not.

If they were similar, with respect to the point of view of the examiner, the machine is classified as an AI (Artificial Intelligence). The test is very reliable if each question could have many possibilities.

2. Human-Computer Interaction (HCI)

according to cognitive psychology studies, a human processes data in a specific way and this test evaluates the interaction between humans and machines. The HCI model is divided into five levels:

- task level
- semantic level
- syntactic level
- interactive level
- a level of physical devices

Then the obtained information is processed by:

- reasoning

- problem solving
- skill acquisition
- error

3. Human Interactive Proof (HIP)

it's used to classify machines and humans looking at a type of interaction, that is simple to be done by a human instead of a bot. The main goals of the test are:

- To differentiate the humans from the computers
- To differentiate a category of the humans
- To differentiate a specific human from the category of humans

HIP consists of a test program administered to humans and computers. The main problem of this proof is that the complexity of the task is very complex and only a specific group of humans can usually solve the test.

The design of CAPTCHA challenges takes inspiration from the previous test because its goal is to establish if the user is a bot or a human. A bot is usually an advanced program that exploits some AI information and pretends to be a human. A CAPTCHA challenge sometimes is so complex that a human user can't solve it. For this reason and in order to guarantee an high level of security, a CAPTCHA has to satisfy the following requirements:

- The solution of a CAPTCHA isn't conditional and it shouldn't depend on the user's language and/or age;
- The solution of a CAPTCHA must be easy for the humans and hard for the bots. In fact humans should solve it in no longer than 30 seconds with very high success rate;
- The creation of a CAPTCHA shouldn't affect the user privacy.

Over the year, the CAPTCHAs must satisfy more and more requirements and the challenges become more and more complex reducing the usability of them. For this reason, modern CAPTCHAs have been developed in the last years. During the design phase, the new challenges are related to a good exploitation of the biometrics and the sensors of modern devices. In addition, machine learning techniques and side-channel information are used to evaluate the user behaviour instead of generating challenges based on cognitive knowledge.

2.1 Traditional CAPTCHAs

The traditional CAPTCHAs are based on the knowledge and the correct insertion of the solution by the user. These CAPTCHA schemes are designed to exploit character recognition, image analysis and speech recognition to guarantee that the challenges will successfully block bots.

The main types of traditional CAPTCHAs are described in the following sections but the details about specific implementations can be found in the article of Walid Khalifa Abdullah Hasan[3]. With respect to user experience, the most enjoyable traditional CAPTCHAs are usually the game-based and image-based ones but the most frustrating CAPTCHA is the text-based one[4]. A summary of usability and security issues, explained in the following sections, is also reported in **Table 2.1**.

2.1.1 Audio-based CAPTCHAs

This type of CAPTCHAs asks the user to type the words contained in an audio file (see **Figure 2.1**). It's developed for vision-impaired users. Audio-based CAPTCHAs usually have problems in usability related to the language dictionary, from which the words are taken, and the similarity of sound for several words. This CAPTCHA is a hard task even for blind users, in fact during an experiment only 46% of the challenges were solved by the participants[5].

One of the most popular Audio-Based CAPTCHAs is *Audio reCAPTCHA*, developed at Carnegie Mellon University and then bought by Google. In this scheme, the user needs to recognize and write a set of 8 spoken characters from a noisy audio file with several background voices. If the user makes a mistake, the test declares that he's a bot.

Audio-based CAPTCHAs are vulnerable to many Automatic Speech Recognition (ASR) programs[6] but also Deep Learning techniques (e.g. DeepCRACK[7]). A good overview about results, obtained by several classification methods, is described in the work of Jennifer Tam et al.[8].

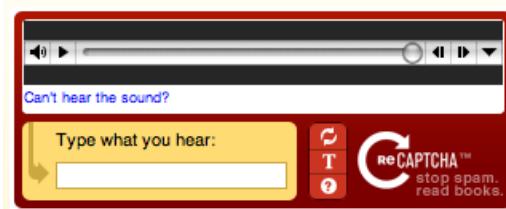


Figure 2.1: Example of audio-based CAPTCHA.

2.1.2 Game-based CAPTCHAs

This type of CAPTCHAs performs the verification of the user identity through a set of several kinds of games (see **Figure 2.2**). The strength of Game-based CAPTCHAs is related to the rules of the game that only humans can understand.

There exists an implementation of these CAPTCHAs, called *Dynamic Cognitive Game (DCG)*, that is usually developed using Flash, HTML5 and JavaScript. These technologies download the game code to the client and execute it locally.

The only difficult for a bot to attack the challenge is the encryption/obfuscation of the code. This strategy prevents the code from being stored on different internet domains. However for example, there exists a bot attack, called *Stream Relay Attack*, that obtains good results bypassing these challenges [9] (see **Section 2.3**).



Figure 2.2: Examples of game-based CAPTCHAs.

2.1.3 Image-based CAPTCHAs

Image-based CAPTCHAs is based on a written text that describes a task related to the evaluation of some images. The user must understand the action and he passes the test if he performs it in the right way. This type

of CAPTCHAs can be categorized into the following classes, looking at the task that the user needs to perform:

- **Click-based CAPTCHAs**

this type of CAPTCHAs shows an image and a text that explains where the user needs to click (see **Figure 2.3**).

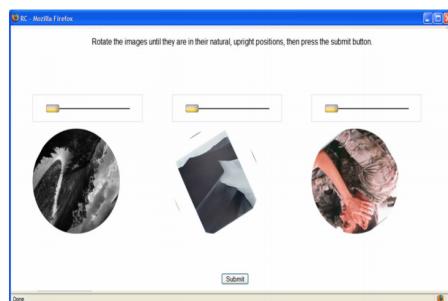
Please click the *circle*, *heart* and *pentagram* regions with different styles:



Figure 2.3: Example of click-based CAPTCHA.

- **Sliding image-based CAPTCHAs**

this type of CAPTCHAs asks the user to use the slider to solve an image-based challenge such as adjusting the orientation of an image, selecting the correct form of an image, or moving a fragment of an image to the correct location (see **Figure 2.4**).



(a) Orientation based.



(b) Form based.

Figure 2.4: Examples of sliding image-based CAPTCHAs.

- **Drag & Drop-based CAPTCHAs**

this type of CAPTCHAs usually asks the user to complete a visual puzzle, created by dividing a given image in a set of pieces[10] (see **Figure 2.5a**).

The task isn't easy for users because this type of CAPTCHAs takes more time to solve the puzzle but the security level is very high[10]. To improve the usability of the CAPTCHA, there exists a version of the puzzle-based CAPTCHA in which the user needs to insert only some pieces of the puzzle instead to complete it (see **Figure 2.5b**).

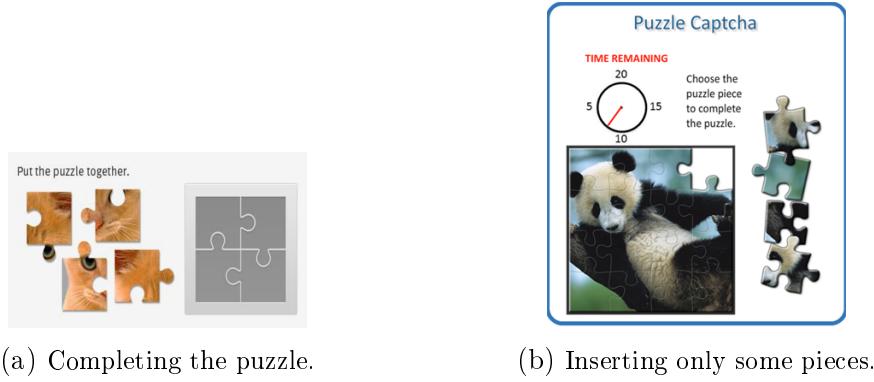


Figure 2.5: Examples of puzzle-based CAPTCHAs.

- **Selection-based CAPTCHAs**

the user usually needs to select the images that contain a requested subject. The set of images, on which the user needs to identify the subject, can be created in several ways, for example:

- An image is divided into a set of sub-squares and each of them is a candidate image (see **Figure 2.6a**)
- There are many images, each one with a unique different subject (see **Figure 2.6b**)

This type of CAPTCHAs is vulnerable to different Object Recognition techniques developed for Computer Vision purposes.

An extension of this type of CAPTCHAs, called *FaceDCAPTCHA*, has been introduced[11] and it incorporates elements of face detection. The human brain is very effective in the process of natural face segmentation even if there are complex backgrounds. In fact, this approach increases the security efficiency because the Computer Vision programs can easily detect if there is a face, e.g. Viola-Jones algorithm[12], but they have many problems understanding if the photographs of the faces are real or not.

One of the most difficult challenges, to be performed by a computer, is the detection of faces, fingerprints and eyes in an image. For this

reason the previous idea has been used to develop a new version of image-based CAPTCHA, called *MB CAPTCHA*[13].

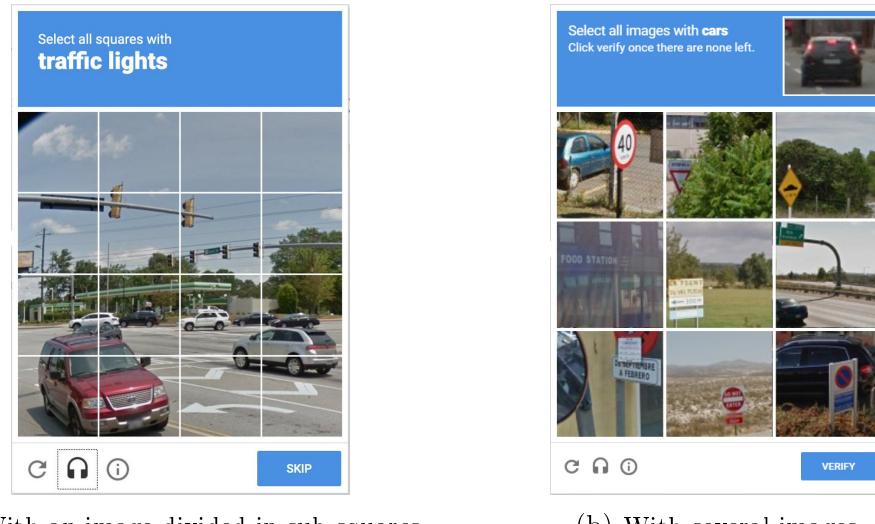


Figure 2.6: Examples of selection-based CAPTCHAs.

- **Interactive-based CAPTCHAs**

the user needs to discover a secret position in an image using mouse movements or swiping gestures

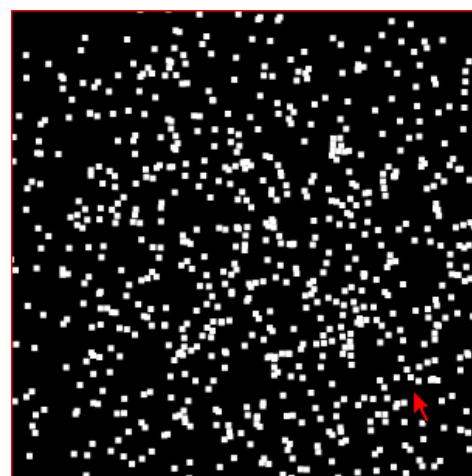


Figure 2.7: Example of interactive-based CAPTCHA.

2.1.4 Math CAPTCHAs

Looking to a math operation, specified in a frame, the user needs to insert the result in a text field. The operation is written in plain text or, to improve the security of this challenge, it's warped like text-based CAPTCHAs (see **Figure 2.8**). These classical math-CAPTCHAs, also known as *arithmetic CAPTCHAs*, are vulnerable to OCR (Optical Character Recognition) techniques.

An advanced version of this CAPTCHA is used in the Quantum Random

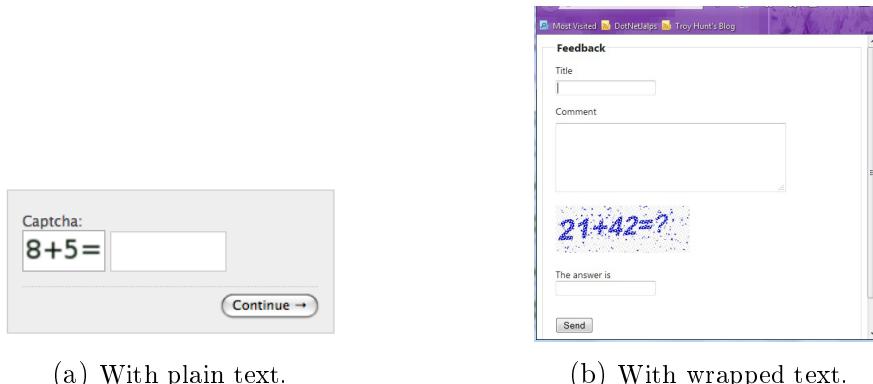


Figure 2.8: Example of arithmetic CAPTCHAs.

Bit Generator Service (QRBGS) sign-up Web Page[14] (see **Figure 2.9**). This type of CAPTCHA asks the user to solve an advanced math expression. It prevents the use of free or commercial OCR techniques because many mathematical symbols are not considered in the traditional detection algorithms. Hence many math symbols are wrongly translated by bot programs and the challenge is very secure. However, this CAPTCHA is vulnerable to side-channel attacks [14] and it's very complex for normal users. In fact many people can't solve it correctly because the required level of math knowledge is very high and not very common.

2.1.5 Slider CAPTCHAs

Slider CAPTCHAs ask users to move a slider across the screen. The image recognition is not part of the challenge.

The most popular CAPTCHAs of this type are the following:

- **CAPTCHA used by Taobao.com**

it asks the user to drag the slider from the start to the end of the sliding bar to verify his identity (see **Figure 2.10a**).

Qualifying question

Just to prove you are a human, please answer the following math challenge.

Q: Calculate:

$$\frac{\partial}{\partial x} \left[5 \cdot \sin \left(7 \cdot x + \frac{\pi}{2} \right) + 2 \cdot \cos \left(5 \cdot x + \frac{\pi}{2} \right) \right] \Big|_{x=0}$$

A: mandatory

Note: If you do not know the answer to this question, reload the page and you'll (probably) get another, easier, question.

Figure 2.9: Example of Quantum Random Bit Generator Service (QRBGS) sign-up Web Page [14].

- **CAPTCHA used by TheyMakeApps.com**

it asks the user to move the slider to the end of the line to submit a form[15] (see **Figure 2.10b**).

Over the years several implementations of this type of CAPTCHAs have been bypassed with a simple JavaScript code and puppeteer.

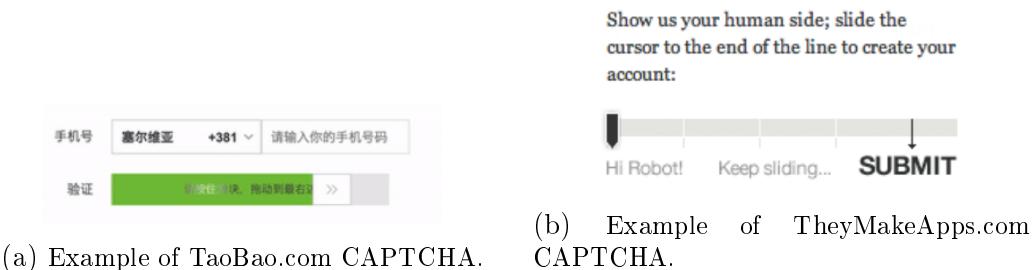


Figure 2.10: Examples of slider CAPTCHAs.

2.1.6 Text-based CAPTCHAs

In text-based CAPTCHA schemes a random series of wrapped characters and/or numbers is displayed on the screen inside an image (see **Figure 2.11**). The user needs to understand which are the characters that composes the sequence and then he types them inside a text-field. The text-based CAPTCHAs can be also classified in three main classes looking to the type of wrapped characters:

- **2D**

the digits are wrapped on a 2D plane, parallel to the screen plane

- **3D**

the digits are wrapped on a 3D plane oriented in the space and then a 2D image is taken from a specific point of view

This type of CAPTCHAs is vulnerable to several attacks, related to Computer Vision techniques:

- OCR techniques[16]
- Segmentation techniques (e.g. DECAPTCHA[17])
- Machine Learning and Deep Learning techniques

In the design phase of a text-based CAPTCHA there are many issues, related to Computer Vision attacks, to be considered. For each of them, there is usually a solution in the design phase of the CAPTCHA that reduces the probability that the challenge will be broken by a bot[17].



Figure 2.11: Example of text-based CAPTCHA.

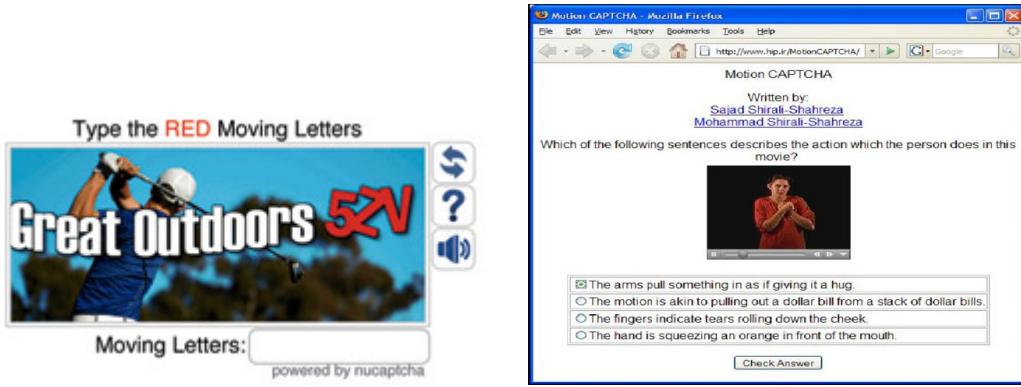
2.1.7 Video-based CAPTCHAs

This type of CAPTCHAs isn't very common because it requires the download of a very huge file[3]. The traditional video-based CAPTCHA is composed by a video in which a sliding text is shown (see **Figure 2.12a**). The user needs to type this message in a text field to pass the challenge. Some implementations of this type of CAPTCHAs are vulnerable to machine learning attacks.

Another version of the CAPTCHA is *Motion CAPTCHA*[18], developed by M. Shirali-shahreza and S. Shirali-shahreza, in which the user needs to watch a video. Then he needs to select which action was performed in the played file, choosing it from multiple answers (see **Figure 2.12b**). The strength of these implementations of CAPTCHAs is in the relationship between the

multiple choices of the answer[19].

A similar implementation of the previous version is the one developed by Kluever et al. in which the user watches a video and then he needs to write three words that describe what he sees[20]. The same authors created a tag frequency-based attack to evaluate the security of their CAPTCHA scheme and they achieve a success rate of 13%.



(a) Example of sliding text in video.

(b) Example of Motion CAPTCHA[18].

Figure 2.12: Examples of video-based CAPTCHAs.

2.2 Modern CAPTCHAs

The type of CAPTCHAs and authentication mechanisms, described in the following sections, are far from traditional CAPTCHAs because they aren't based on cognitive knowledge of a human user. They work exploiting other parameters, such as behavioural analysis and data of the sensors. In the following sections there is a summary of the most known CAPTCHA schemes of this type.

2.2.1 Biometrics-based CAPTCHAs

Biometric-based CAPTCHAs are very used to support the authentication mechanisms of a user. The most known schemes of this type are the following:

- **Bio-CAPTCHA voice-based Authentication**

This authentication method was developed starting from good results reached in the authentication phase of cloud systems[21] (Alexa for Amazon, Siri for Apple, Cortana for Microsoft). This particular implementation uses a random voice-based password challenge. This password changes at every login and the user needs to say it out loud. The

strength of the challenge is the peculiarity of the human voice. The experiments reveals that unauthorized access probability decreases, while the usability is very high because it requires only the access to the microphone.

- **rtCAPTCHA**

this type of authentication method is a Real-time CAPTCHA that asks users to perform some tasks like smile, blink or nod in front of the camera of the mobile phone. The recorded video is sent to the service provider that checks if there is the expected user performing the required action in the file.

This implementation of CAPTCHA solves many problems of similar CAPTCHAs that are also based on liveness mechanisms and video capture. An attacker can extract patterns or features from existing or captured images of the user. Then he embeds them into a new generated video to attack the authentication system in the compromised device.

In the work of Erkam Uzun, Simon Pak Ho Chung, Irfan Essa and Wenke Lee[22], there is a detailed comparison between other similar authentication mechanisms and rtCAPTCHA, looking to all possible Computer Vision attacks.

2.2.2 Behavioural-based CAPTCHAs

In 2014 Google announced that the Artificial Intelligence can solve now even the most difficult variant of text-based CAPTCHAs at 99.8% accuracy[23]. For this reason, the company develops the following CAPTCHA schemes:

- **Google no CAPTCHA**

Google developed this new CAPTCHA system in 2015. It's simpler than traditional CAPTCHAs in terms of user interaction[24]. This CAPTCHA system is composed by two layers of protection:

1. Checkbox "*I'm not a robot*" to be clicked by the user as shown in **Figure 2.13** (or image-based CAPTCHA on mobile devices)

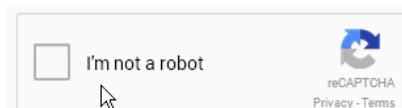


Figure 2.13: Example of Google no CAPTCHA checkbox.

2. Traditional text-based CAPTCHA with two warped words

The second layer is reached only if the user doesn't succeed in the first one. For the checkbox step, the application evaluates in background the user's behaviour (e.g. the mouse movement, where the user clicks, how long he lingers over a checkbox). Then the program performs an *advanced risk analysis*, by looking at the results of the first step, the spam traffic and the number of passed/failed challenges. In this way the CAPTCHA understands if the test is passed or not.

The experimental results confirm that the first phase is very inefficient and many times the user fails performing this challenge even if he was a human and was performing correctly the task. A problem of this type of CAPTCHA is that many attacks exploits the image-based CAPTCHA (used on mobile phones) and the text-based CAPTCHA of the second phase using attacks based on known Computer Vision techniques or their variants (e.g. CAPTCHA breaker made by Suphanee Sivakorn, Jason Polakis and Angelos D. Keromytis[25]).

• Google Invisible ReCAPTCHA

it's a top layer over the *Google noCAPTCHA v2.0*. It adds the option to bind the CAPTCHA directly to the form's submit element. In this way, the programmer can add layers to reduce the user experience of a bot[24]. It usually requires the use of cookies to track the user's behaviour. There exist two version of this CAPTCHA:

– ReCAPTCHA v2.0

it was developed in 2017. It's not really invisible because Privacy & Policy badge must be included on every pages of an app or a website in which the CAPTCHA is used. Computer Vision and Artificial Intelligence algorithms can break the challenges by recognizing object in the pictures in the image-based CAPTCHA phase.

– ReCAPTCHA v3.0

it was developed in 2018. It constantly analyses the human behavior, the mouse movements, the typing speed and the other features in NO CAPTCHA technology until Google collects enough data to tune their Google Invisible reCAPTCHA v2.0. This type of CAPTCHAs uses the probability scores of Artificial Intelligence and Machine Learning, the hostname, the timestamp and the action validation.

Google removes the image recognition phase and looking only at

the score, it evaluates if the user is a human or a bot. The main difference with respect to the previous versions of Google modern CAPTCHAs is that it returns a probability score, called *risk score*, in the range $[0.0, 1.0]$: *0.0* if the user is a bot, *1.0* otherwise. The administrator of the website can decide what range of scores he wants to manage, declaring when the site is under attack and what actions need to be performed.

Over the year, several peculiarities and problems of this version of *Invisible ReCAPTCHA* were discovered. For example:

- * If a user accesses a Web page using incognito mode or private mode, he is classified with a very low score (*high risk*).
- * If a human is wrongly classified as a bot, the user can login into its Google account to increase its score. If this doesn't change the classification, you cannot do anything else.

2.2.3 Sensor-based CAPTCHAs

The devices, with access to the net, have natively many sensors nowadays, like the gyroscope and the accelerometer. The CAPTCHA schemes, described in the following sections, exploit the data generated from these sensors to improve the security of the authentication.

- **Completely Automated Public Physical test to tell Computer and Humans Apart (CAPPCHA)**
this is a way to enforce the PIN authentication phase by mobile phone[26].
The user needs to tilt the device to a specified angle, specified on the screen (see **Figure 2.14**). The CAPPCHA security is based on the *Secure Element (SE)* in the device. It prevents brute force, side channel and recording attacks. The usability results are good because many comments, made by the users, were also considered in the implementation.



Figure 2.14: CAPPCHA and PIN authentication[26].

- **Invisible CAPTCHA**

It will be described in details in **Chapter 4**.

2.3 CAPTCHA security

The process used to break many traditional CAPTCHAs, based on text or images, is usually organized into the following phases[27]:

1. **Pre-processing phase**

in this phase, several techniques are applied to remove background, to separate foreground from the background, to delete noise and to remove some particular pattern (e.g. Canny Detection and Scale-Invariant Feature Transform (SIFT) application).

2. **Attack phase**

the following techniques are usually applied:

- ***Object Segmentation attacks***

Segmentation techniques (e.g. vertical histogram, colorfilling, snake segmentation and JSEG) are used to split the CAPTCHA image into segments to simplify the recognition

- ***Object recognition attacks***

The most used techniques are pattern matching (e.g. shape context matching, correlation algorithm), OCR recognition, SIFT and machine learning.

- ***Random Guess Attacks***

The attacker's program tries to break the CAPTCHA scheme by guessing the correct answer. This attack is effective on CAPTCHAs with few number of different challenges.

- ***Human Solver Relay Attacks***

The bot forwards the CAPTCHA challenge to a remote human worker that will solve it if the previous phases don't produce good results. This technique is usually used also to solve other types of CAPTCHAs.

Many CAPTCHA schemes have still several issues:

- **Session issue**

Some types of CAPTCHAs have a big issue because they don't destroy the session, after a correct answer is inserted by the user[1].

Hence, the hacker can crack the next 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.

- **Resilience to both automated and human solver relay attacks**

Many CAPTCHA schemes are designed to be robust against a possible AI attack but the new generation of CAPTCHAs involves the use of remote bot or human solver. Traditional CAPTCHA schemes are vulnerable to this type of attacks.

Invisible reCAPTCHA and other academic proposals haven't been attacked yet, but works over thousands of different IP addresses and simulate the human behaviour. Sensor-based CAPTCHAs are often vulnerable to relay attacks. An exception is the *Invisible CAPPCHA*, that will be analysed in **Chapter 4**.

- **Limited number of challenges**

An issue of sensor-based CAPTCHA schemes is the limited number of challenges because the design of many usable gestures is very hard. This problem could be solved relying on trusted hardware.

- **Trade-off between Friction-heavy and Frictionless CAPTCHAs**

A trade-off between usability and security aspects is always considered analysing CAPTCHA schemes. This condition is highlighted in behavioural and sensor-based CAPTCHAs.

- **User's privacy**

Sensor-based and behavioural CAPTCHAs usually send useful information to a remote server that analyses it to establish if the user was a human or a bot. If an hacker attacks the server side of this application, he can access to users' private data.

In some CAPTCHAs, the information is evaluated on the client side by a trusted hardware and the server receives only the results of the analysis. In this case, we need to be sure that trusted hardware is secure enough to guarantee privacy of user's information.

- **Compatibility with different devices**

Many CAPTCHA schemes, e.g. behavioural ones, use specific forms factors but a good challenge should be compatible with different factors.

Type	Scheme	Usability issues	Security
Audio	<i>Audio reCAPTCHA</i>	Issues of recognition: <ul style="list-style-type: none"> • Knowledge of English dictionary by the user. • Some character sounds very similar to others. 	It's vulnerable to: <ul style="list-style-type: none"> • ASR programs. • Deep Learning and ML techniques.
	<i>Dynamic Cognitive Game (DCG)</i>	Comprehension of rules.	Vulnerable to Stream Relay Attack
Image	<i>Click-based Drag & Drop-based Sliding-based Selection-based Interactive-based</i>	Difficulty in identification of images caused by: <ul style="list-style-type: none"> • Blur of images. • Low vision condition. 	Vulnerable to: <ul style="list-style-type: none"> • Segmentation techniques • Deep Learning and ML techniques • OCR techniques
	<i>Arithmetic QRBG Selection-based</i>	It requires basic or advanced math knowledge.	Vulnerable to: <ul style="list-style-type: none"> • OCR techniques • Side-channel attacks
	<i>Slider TheyMakeApps.com</i>	Simple and intuitive interaction	Simple bypassed by Javascript code and pupeeteer
	<i>2D 3D</i>	Many problems have to be solved by user: <ul style="list-style-type: none"> • Multiple fonts • Font size • Blurred Letters • Wave Motion 	It can be identified by: <ul style="list-style-type: none"> • OCR technique • Segmentation techniques • Deep Learning and ML techniques
Video	<i>Motion CAPTCHA</i>	Heavy file to be downloaded	

Table 2.1: Survey of main types of traditional CAPTCHAs[10].

Chapter 3

Side-channel attacks

A side-channel attack is an attack in which the malicious user exploits a side-information of transmitted encrypted data, to give access to the user's private data. This type of extra information is usually: timing information, power consumption, electromagnetic radiations, sound and so on.

In the past the main side-channel attacks require the physical access to the victim's device. Nowadays, side-channel attacks are evolved and can be conducted by remote hackers using malicious code (e.g. cache-timing attacks, DRAM row buffer attacks), even exploiting information from sensors on mobile devices[28].

In fact an hacker can obtain side-channel information on mobile devices, from several sensors[29]:

- Location sensors (e.g. GPS, proximity)
- Motion sensors (e.g. accelerometer, gyroscope, magnetometer)
- Environmental sensors (e.g. for ambient light, temperature, barometer)
- Biometric sensors for wearable devices (e.g. heart rate sensor, ECG)
- Audio sensors (microphone)
- Video sensors (camera)

Side-channel attacks can usually be classified, looking at the action performed by the attacker, in the following categories:

- **active**
the hacker influences the behaviour of the victim's device
- **passive**
the attacker only analyses the leaking information

Side-channel attacks can also be classified, looking at the distance between the hacker and the victim during the attack, in the following classes:

- Local attacks
- Vicinity attacks
- Remote attacks

In this chapter there is a summary of the most popular side-channel attacks, organized with respect to the previous classifications (see **Table 3.2**). This study was conducted to find a feasible side-channel to be used in the design phase of a new CAPTCHA scheme, as support to the authentication phase.

3.1 Local side-channel attacks

In local attacks, the attacker needs to get the target device or to be very near to it. In many cases the hacker physically needs to manipulate the device or to obtain access to the chip.

Passive attacks

The following attacks are used to break cryptographic system implementations:

- ***Power Analysis***

this type of attacks are based on the analysis of the power variations in transistors. There exist several popular attacks[30]:

- ***Simple Power Analysis (SPA)***

the attacker analyses the power consumption of the system, that depends on the microprocessor used. This analysis can be useful to understand which operations are performed by different implementations of cryptographic algorithm (e.g. RSA, DES).

- ***Differential Power Analysis (DPA)***

this attack collects data and then performs a statistical analysis and some error correction techniques from data to extract information correlated with secret keys.

- ***High Order DPA (HO-DPA)***

While DPA obtains information across a single event, HO-DPA correlates with multiple cryptographic sub-operations.

- ***Electromagnetic Analysis Attacks***

the attacker can analyse indirectly the power consumption by accessing electromagnetic signals of the victim's machine. This type of attacks depends on the used instruments (e.g. EM probes) and on the analysed location of the chip. These dependencies also affect the signal-to-noise ratio.

- ***Differential Computation Analysis***

the attacker tries to exploit white-box cryptographic implementations. In this model the attacker, even if he has access to code, can't extract the secret key. Hence the hacker needs to have full control over the target device and the execution environment. Then using binary instruments, he can control the intermediate states of the memory or memory operations (e.g. reading/writing operations)[31].

- ***Smudge Attacks***

the attacker can exploit fingerprints and smudges on the screen of mobile devices to evaluate the user's input.

- ***Shoulder Surfing and Reflections***

the attacker can exploit the lightness of the device display and obtain the user's activity by its reflection on sunglasses or tea pods.

- ***Hand/Device Movements***

the attacker exploits the user's movements of fingers and hand to understand the interaction of the victim with the device.

Active attacks

The following attacks require that the hacker physically gets the device for a while:

- ***Clock/Power Glitching***

in the past the attacker can fault inject on embedded devices by exploiting variations of the clock signal, (e.g. overclocking). However he needs to use an external clock source to do it.

- ***Electromagnetic Fault Injection (EMFI)***

the attacker uses short (e.g. nanoseconds) high-energy electromagnetic pulses to change the state of memory cells. This attack allows to target specific regions of a microchip by locating the EM probe (e.g. on the instruction memory, the data memory, or CPU registers).

- ***Laser/Optical Faults***

the attacker needs to decapsulate the chip to obtain access to it and,

using a laser beam, it change the state of a transistor (e.g. changing bit value of a memory cell).

- ***Temperature Variation***

the attacker can change the temperature, in which the target device is working, causing malfunctioning of the hardware. Temperature higher than the maximum one, supported by the hardware, causes faults in memory cells. Temperature too low changes the speed, for which the content of the RAM disappear, after turning the device off.

- ***Differential Computation Analysis***

the attacker needs to have full control of the white-box environment, manipulating intermediate values in the system computation.

- ***NAND Mirroring***

the attacker exploits the duplication of the data, usually used to recover data after faults, to restore a previous system state. The hacker can force the reset of the state as demonstrated by Skorobogatov for an Apple device[32].

3.2 Vicinity side-channel attacks

The attacker needs to wiretap or eavesdrop the network communication of the victim or to be in the neighbourhood of the target.

Passive attacks

- ***Network Traffic Analysis***

the attacker can exploit meta data, related to the encrypted data, transmitted over the network. This information gives access to sensitive information about the traffic.

For example a Web-application works between two parties: the client and the server. For this reason the communication channel is usually encrypted and the requests made by the user work through the *HTTPS* protocol.

This solution isn't enough to prevent an attacker from exploiting reserved data because each web page has a distinct size and loads resources of different sizes. Hence the attacker can fingerprint the page even if HTTPS protocol is used.

Another cause of these attacks on Web-services is given by the trend of Web to work on *stateful protocols*, providing better performance to

the client by keeping track of the connection information. TCP session for example works on Stateful Protocol because both systems maintain information about the session itself during its life[33].

- ***USB Power Analysis***

the attacker can modify USB charging stations for mobile devices to obtain an analysis about the power consumption and the related sensitive information.

- ***Wi-Fi Signal Monitoring***

Wi-Fi devices continuously monitor the wireless channel (channel state information (CSI)) to transmit data. Any environmental variation (e.g. finger motion) affects Wireless signals, generating unique patterns. For example the attacker can exploit these variations to unlock patterns on smartphones[34].

Active attacks

- ***Network Traffic Analysis***

the attacker, after obtaining information about transmitted packets, can interfere the traffic (e.g. delay of packets).

3.3 Remote side-channel attacks

These attacks are software-only based and they depend on the installation of the malicious code on the target device.

Passive attacks

- ***Linux-inherited procfs Leaks***

the attacker can obtain a lot of information for each running process on Linux File System, looking at the content of the files in **Table 3.1**.

File path	Information
<code>/proc/[pid]/statm</code>	Virtual and physical memory sizes of process with identifier <code>[pid]</code>
<code>/proc/[pid]/stat</code>	CPU utilization times of process with identifier <code>[pid]</code>
<code>/proc/[pid]/status</code>	Number of context switches of process with identifier <code>[pid]</code>
<code>/proc/interrupts</code>	Interrupt counters
<code>/proc/stat</code>	Context switches

Table 3.1: Files used to obtain information about running processes.

- ***Data-Usage Statistics***

the attacker can access to information, about incoming and outgoing network traffic, of each application without any permission.

- ***Page Deduplication***

To reduce the overall memory footprint of a system, some operating systems perform deduplication, searching for identical pages within the physical memory and merge them even across different processes. When a process tries to write on a deduplicated page, a copy-on-write fault occurs and the process gets its own copy of this memory region again.

- ***Microarchitectural Attacks***

By measuring execution times and memory accesses, the attacker can obtain sensitive information from processes running in parallel on the same device. This type of information can be evaluated from CPU caches, that are a big source of information leaks.

- ***Sensor-based Keyloggers***

in mobile devices, the attacker can exploit information from equipped sensors with any permission. The user's interaction with the device can be evaluated by analysing information from the sensors.

- ***Fingerprinting Devices/Users***

the attacker can obtain the identity of the device and the user and also the fingerprint by exploiting hardware issues and cookies.

- ***Location Inference***

the attacker can obtain user's location without using GPS sensor, that requires permission to be accessed. For example, the accelerometer and the gyroscope can be used to infer car driving routes.

- ***Speech Recognition***

the access to the microphone is protected by permissions. The attacker can also exploit gyroscope to obtain information about human speech near to the device.

- ***Soundcomber***

if the attacker obtains permission of the microphone, he can obtain sensitive information (e.g. credit card numbers) on the automated menu services of phones.

Active attacks

- ***Rowhammer***

the attacker can induce hardware faults by frequent accesses to main memory. This happens because the size of DRAM cells decreases nowadays to increase the density of memory cells in DRAM. However this causes electromagnetic coupling effects between near cells.

	Local	Vicinity	Remote
Passive	<ul style="list-style-type: none"> • Power Analysis • Electromagnetic Analysis Attacks • Differential Computation Analysis • Smudge Attacks • Shoulder Surfing and Reflections • Hand/Device Movements 	<ul style="list-style-type: none"> • Network Traffic Analysis • USB Power Analysis • Wi-Fi Signal Monitoring • Sensor-based Keyloggers • Fingerprinting Devices/Users • Location Inference • Speech Recognition • Soundcomber 	<ul style="list-style-type: none"> • Linux-inherited procfs Leaks • Data-Usage Statistics • Page Deduplication • Microarchitectural Attacks • Sensor-based Keyloggers • Fingerprinting Devices/Users • Location Inference • Speech Recognition • Soundcomber
Active	<ul style="list-style-type: none"> • Clock/Power Glitching • Electromagnetic Fault Injection (EMFI) • Laser/Optical Faults • Temperature Variation • Differential Computation Analysis • NAND Mirroring 	<ul style="list-style-type: none"> • Network Traffic Analysis 	<ul style="list-style-type: none"> • Rowhammer

Table 3.2: Survey of the most popular side-channel attacks[28].

Chapter 4

Invisible CAPPCHA

The *Invisible CAPPCHA* is an evolution of CAPPCHA, in terms of usability[35]. The main difference with respect to CAPPCHA is that the challenge isn't explicitly submitted to the user but it's hidden behind the PIN authentication phase. This type of challenge works only on smartphones as its ancestor but it's very important to understand how data, obtained by the sensors, could be exploited during the authentication phase.

Invisible CAPPCHA is a method developed in 2018 and based on motion side-channel of mobile devices and it's very effective as support of Password-based authentication methods. The main steps, followed by this CAPTCHA, are:

1. **Motion detection**

in this phase the micro-movements of the device, generated by the interaction of the user with the touch-screen, are evaluated by the *Secure Element (SE)*

2. **Communication between Client and Server**

in this phase the credentials are shared with the remote Service Provider

The following sections will analyse this steps highlighting some peculiarities of Invisible CAPPCHA that will be used or modified to develop AcCAP-PCHA, a CAPTCHA based on microphone sensors.

4.1 Motion detection

In this first phase, Invisible CAPPCHA exploits the accelerometer of the mobile device. It detects the acceleration over the three axis in g-force units, as a sequence of vectors over time:

$$\{A_i\}_{i=1}^n = \{(a_1^x, a_1^y, a_1^z), \dots, (a_n^x, a_n^y, a_n^z)\}$$

This type of side-channel information from embedded accelerometer has been exploited in different attacks for the single and double tap detection. These attacks analyse the accelerations over the z-axis by comparing them to some thresholds and some timing conditions.

In Invisible CAPPCHA, the side-channel information is stored on the memory of the mobile device. Depending on the device, a smartphone built-in vibration can be generated along Z axis or along more than one axis (see **Figure 4.1**). At the contrary a finger tap event is defined by strong accelerations on Z-axis but also similar one to the other (see **Figure 4.2**).

In Invisible CAPPCHA, the difference between built-in vibration and tap acceleration is evaluated by a simple algorithm. It relies on negative and positive peaks, that are detected by comparing acceleration along Z axis with predefined thresholds. The differences between the tap accelerations and the vibrations is the main characteristic that guarantees user's tap cannot be simulated by a bot using the vibration motors.

Another important requirement, to prevent a bot attack, is that Invisible CAPPCHA uses a Secure Element that embedded the accelerometer of the device. In this way malicious code can't access to the sensor. Nowadays there exists a smart card, called SIMSense, that already integrates motion sensor and embeds it in a Secure Element.

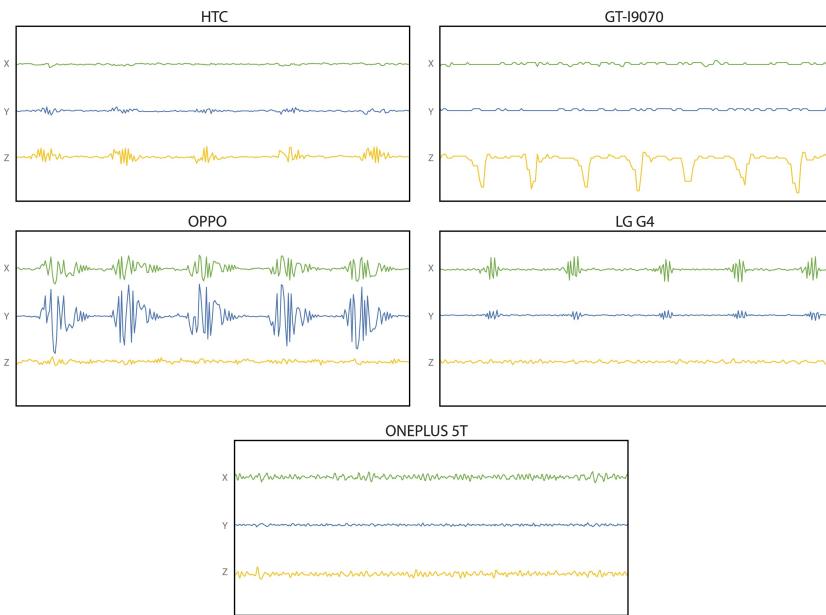


Figure 4.1: Example of accelerations caused by smartphone built-in vibration.

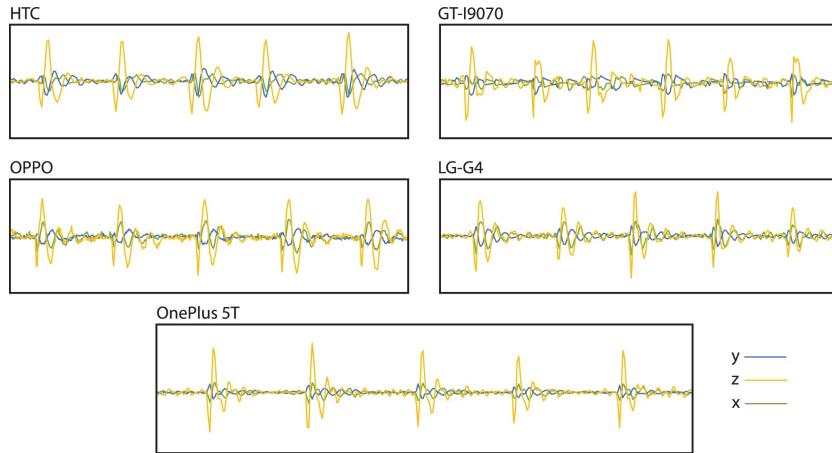


Figure 4.2: Example of accelerations caused by finger tap detection.

4.2 Communication between Client and Server

When the user fills a form or provides other information to a cloud application/service, the Secure Element checks if a micro-movement is measured when a user tap is detected. If this happens the input inserted by user is considered valid, or rather generated by a human, otherwise the algorithm tells that the input was generated by a bot.

An extra message, that tells if the task was performed by a user or not, is sent to the server side. The integrity of this message is guaranteed by the Secure Element, that can be equipped with a digital signature. The identity of the device can be associated to the message sent and then it can be checked and verified. The Secure Element signs the verification message through ECDSA.

4.2.1 Elliptic Curve Digital Signature Algorithm (ECDSA)

The elliptic cryptography works similarly to RSA but it uses smaller keys. The signature algorithm with elliptic curves is divided in two phases, like the one based on RSA. Considering two users, Alice and Bob, ECDSA phases are explained in the following lines:

- **Sign generation**

If Alice wants to send a message, protected with digital sign, to Bob, they need to share the following parameters (*curve*, *G*, *n*): *curve* is the equation of an elliptic curve, *G* is the base point of prime order on the curve and *n* is the multiplicative order of *G* for which $n \times G = O$ (*x* is the scalar multiplication of a point of the curve). Alice generates a

private key d_A in the range $[1, n - 1]$ and a public key $Q_A = d_A \times G$. Alice needs to perform **Algorithm 1** to sign a message m .

Algorithm 1: Sign generation.

Input: m = message to be signed

Output: (r, s) = digital sign

```

1  $e \leftarrow \text{HASH}(m)$  where  $\text{HASH}$  is an hash function (e.g. SHA-2)
2  $z \leftarrow$  string composed by the  $L_n$  most left bits
3     where  $L_n$  is the bit length of the group of order  $n$ 
4  $r \leftarrow 0$ 
5  $s \leftarrow 0$ 
6 while  $r = 0 \bmod n$  or  $r = 0 \bmod n$  do
7      $k \leftarrow \text{RANDOM}([1, n - 1])$ 
8      $(x_1, y_1) = k \times G$  of the elliptic curve
9      $r \leftarrow x_1 \bmod n$ 
10     $s \leftarrow k^{-1}(z + rd_A) \bmod n$ 
```

- **Sign verification**

Bob wants to verify the digital signature sent by Alice. To do it, he needs to apply in order **Algorithm 2** and **3**.

Algorithm 2: Verification that public key is on the elliptic curve.

Input: Q_A = public key to be verified

Output: check = true if public key is correct

```

1  $\text{check} \leftarrow \text{true}$ 
2 \\Valid coordinates
3 if  $Q_A = O$  then
4      $\text{check} \leftarrow \text{false}$ 
5 \\Element of the curve
6 if  $Q_A \in \text{curve}$  then
7      $\text{check} \leftarrow \text{false}$ 
8 \\Correctness of order
9 if not  $n \times Q_A = O$  then
10     $\text{check} \leftarrow \text{false}$ 
```

In Invisible CAPPCHA the message m is bitwise concatenated with a signed unique value, the nonce n , and then the signature is computed on their concatenation, $m||n$. Hence the signed message sent to the server is (r, s, m, n) .

Algorithm 3: Sign verification.

Input: (r, s, m) = digital sign and message
Output: m = message to be signed

```

1  $e \leftarrow HASH(m)$  where  $HASH$  is an hash function (e.g. SHA-2)
2  $z \leftarrow$  string composed by the  $L_n$  most left bits
3 where  $L_n$  is the bit length of the group of order  $n$ 
4 if  $not r \in [1, n - 1]$  or  $not s \in [1, n - 1]$  then
5    $\quad$  *Invalid sign*
6    $e \leftarrow HASH(m)$ 
7    $z \leftarrow$  string composed by the  $L_n$  most left bits
8    $w = s^{-1} \bmod n$ 
9    $u_1 = zw \bmod n$ 
10   $u_2 = rw \bmod n$ 
11   $(x_1, y_1) = u_1 x G + u_2 x Q_A$  of the elliptic curve
12  if  $r \equiv x_1 \pmod{n}$  then
13     $\quad$  *Verified sign*
14  else
15     $\quad$  *Not accepted sign*
```

4.3 Security analysis

After the verification of the signature, the communication must be also encrypted to ensure integrity and authenticity of exchanged messages. The Secure Element can be accessed only through PIN authentication of the off-card communication party. If the malicious code has enough privileges to access Secure Element, some popular attacks can't be performed by an attacker.

4.3.1 Strength against popular attacks

The most popular attacks, that have been analysed, are[35]:

- **Replay attack**

Because the message is signed together with a nonce, an attacker can't easily use a message already sent by a client to the server. In fact, the server checks if a nonce was already used by the client and if so, the server refuse the message sent by the attacker.

- **Reverse engineering attack**

Even if the attacker can de-obfuscate the code of the application run-

ning on the browser, he can access to reserved data on the server only if the verification message for human interaction was correctly signed by the Secure Element. Hence this type of attack can't be performed.

- **Human-solver relay attack**

The Invisible CAPPCHA is strong to this type of attack because it doesn't require any additional task to be sent to a remote human solver, as in standard CAPTCHAs.

- **Brute force and password replay attacks**

Invisible CAPPCHA can be used to validate every input before it considers it as a possible attempt for a password. If the password was inserted by a malware or was wrong, the number of attempts decreases. Hence this approach prevents a brute force attack. This also prevents the access to the Secure Element by the attacker in replay attacks.

- **Denial Of Service (DOS)**

If a malware tried more than the maximum amount of attempts of passwords it could do a Denial Of Service (DOS) of the Secure Element. To prevent this attack, the Secure Element can block access to itself if three invalid passwords are inserted by a human or if an invalid password is inserted by a bot.

Chapter 5

AcCAPPCHA

Invisible CAPPCHA works exploiting information obtained by a sensor. Starting from this idea and the sensor-based keyloggers in [Section 3.3](#), I decide to design a CAPTCHA that works more or less like a keylogger. A keylogger is usually a malicious program, used to acquire information about the user's activity by exploiting side-channel information. Its implementations depends on the party, that the hacker wants to attack[\[36\]](#):

- **The user**

these attacks are based on the exploitation of physical information related to the typing state. For example, they can use electroencephalography (EEG), motion of the wrist in the smartwatches, video with keyboard line-of-sight and WiFi signal distortion.

- **The keyboard**

these attacks are based on analysis of signals coming from the keyboard. For example, acoustic emanations can be exploited by using external physical sensors.

- **The host**

these attacks are based on the physical access of the attacker to the victim machine. For example, the process footprint, the CPU load and other micro-architectural analysis can be exploited in these attacks.

- **The network**

these attacks exploit the packets exchanged in the client-server communication. For example, a network packet can be related to a keystroke revealing the key press time of the victim and the payload size of the server response.

I develop a new CAPTCHA based on a keylogger that analyses the keyboard. Considering also the Bio-CAPTCHA voice-based Authentication in **Section 2.2.1**, I decide to exploit the audio signal of the microphone of the device. The developed CAPTCHA is called AcCAPPCHA that stands for "Acoustic CAPPCHA". It works in the background of the Authentication phase, like Invisible CAPPCHA, and exploits the acoustic side-channel of the device microphone. This sensor was chosen to simply the usability of a CAPPCHA, because it requires only the access permissions to it, without affecting the strength of its bot detection.

The whole implementation was made using **Python** language. The structure and the behaviour of AcCAPPCHA are similar to the ones proposed in Invisible CAPPCHA because they are both based on the analysis of some signal, detected by some sensor, during the insertion of the password.

The main difference is that the signal analysed by AcCAPPCHA is generated by the microphone, instead of the accelerometer. However AcCAPPCHA introduces also another possible analysis of the audio signals using neural networks.

According to the sequence of actions performed by Invisible CAPPCHA, AcCAPPCHA performs two steps for the authentication of a user:

- Evaluation of the user's activity
- Communication with the remote service

A description of them is reported in the following sections.

5.1 Evaluation of the user's activity

During the evaluation of the user's activity, AcCAPPCHA records two audio signals: the first one created during the insertion of the password and the second one created before. The first signal will be analysed to find the audio peaks. The second signal is exploited to define a threshold, that will be used to find significant audio sequences in the first signal. This procedure was introduced to manage background noise during the classification of the audio. AcCAPPCHA performs two types of verification of the user identity:

- **Time correspondence**

it always requires that the time instants, in which all the characters of the password were typed by the user, must be stored. Then an algorithm checks if there exists a sequence of time peaks, observed in the first audio signal, that overlaps with stored time instants. If the

sequence exists, the insertion was performed by a human, otherwise by a bot.

- **Character correspondence**

it doesn't require to store the time instants as in the previous case. Because it analysis only the first audio signal by labeling each audio peak with a key of the keyboard. Hence AcCAPPCHA checks if there exists a sequence of labels, ordered by increasing time instants, that is equal to the password sequence of characters. If the sequence exists, the insertion was performed by a human, otherwise by a bot. The character correspondence could be also applied after applying a time correspondence. In this case, the only differences are:

- AcCAPPCHA needs to store the time, in which all the characters of the password were typed by the user, to perform time correspondence
- Time correspondence must return the audio peaks that overlap with the stored time instants
- If the time correspondence had positive response, the character correspondence will be applied only of overlapping audio peaks. Otherwise, AcCAPPCHA tells that the insertion was performed by a bot.

This last combination of the two verification approaches is more accurate and strong against false positive (bot detected as human).

In the following sections the previous techniques will be explained in details to understand better their pros and cons.

5.1.1 Time correspondence

The first step of the algorithm is the definition of the noise threshold and it's performed before the application asks the user to type the password and after the insertion of the username. During this phase AcCAPPCHA records an audio file of 2 second, called **noise signal**, from the built-in microphone of the laptop and analyses it. The program looks for the maximum value of the audio signal, called $thresh_N$, that will be used later during the thresholding phase.

The next steps of the verification are performed by two threads simultaneously, during the password insertion. The first thread manages the insertion of the password and it's continuously waiting for the insertion of a character

by the user until CARRIAGE RETURN ('\r') is typed. Immediately after a key is pressed, the time instant of this action, related to the Epoch of the PC, is stored by AcCAPPCHA. The sequence of time instants stored by the thread is called $x = (x_0, \dots, x_{|\text{password}|-1})$.

Meanwhile the second thread records an audio signal, called **user activity**, using the built-in microphone. The recording session begins immediately after the insertion of the username and it ends after the detection of the CARRIAGE RETURN by the first thread. AcCAPPCHA removes the last 200 ms of this audio signal to be sure that the *CARRIAGE RETURN* peak isn't included in the final audio file.

From now on, the application has everything it needs to understand if the user is a human or not. In fact the verification is performed by looking if there exists a sequence of audio peaks in the signal that overlaps with the series of the time instants stored by the first thread.

First of all, AcCAPPCHA performs the thresholding phase: **user activity** is analysed by keeping only the samples with values higher than thresh_N . All the previous operations are performed also in the verification of the character correspondence (see **Section 5.1.2**). The samples, survived after the thresholding, will be grouped in several disjoint windows of maximum width equal to 100 ms. For each group W_i , the application looks for the sample with the highest value, peak_i . For example, given the *sampling period or interval* t_s and a specific group of samples:

$$W_i = (w_t, w_{t+t_s}, \dots, w_{t+\lceil \frac{5ms}{t_s} \rceil * t_s})$$

the application computes $t_i = \text{argmax}(W_i)$, the time instant of the sample related to peak_i .

Given the sequence of computed time instants $t = (t_0, t_1, \dots, t_{n-1})$, related to peaks of all the windows, n the number of windows obtained after thresholding and $|\text{password}|$ the size of the password, there is a **time correspondence** if there exists a subset of t , called $t^* = (t^*_0, \dots, t^*_{|\text{password}|-1})$, that matches with the sequence of time instants stored during the password insertion. **Algorithm 4** is used to find a time correspondence and establish if the password was inserted by a human or not. The procedure evaluates the distances between the time instant of the first audio peak and the time instants of the other ones. Then the algorithm evaluates if there exists an overlay of the time instants of the audio peaks and the stored time instants during the password insertion. At the end the threshold threshold , used in **Algorithm 4**, is 10 ms.

Algorithm 4: Time correspondence

Input: $x = (x_0, x_1, \dots, x_{|\text{password}|-1})$ = time instants stored by first thread
 $t = (t_0, t_1, \dots, t_{n-1})$ = time instants relative to peaks of each group
threshold = threshold with respect to stored time instant

Output: **true** if human, **false** otherwise

```

1   $y = (y_0, y_1, \dots, y_{|\text{password}|-1})$  where  $y_i = x_i - x_0$ 
2  if  $n < |\text{password}|$  then
3      //Number of found peaks lower than number of characters of the password
4      return false
5  //Search of subsequence
6  for  $i \leftarrow 0$  to  $n - 1$  do
7      if  $(n - i) < |\text{password}|$  then
8          //Not enough peaks from  $t_i$  on to be analysed to find the time
9          //correspondence
10         return false
11
12
13  // $t_i$  already verified
14   $j \leftarrow i + 1$ 
15   $count \leftarrow 1$ 
16  while  $count < |\text{password}| \wedge j < n$  do
17      if  $(n - j) < (|\text{password}| - count)$  then
18          //Not enough peaks from  $t_i$  on to be analysed to find the time
19          //correspondence
20          break
21
22      if  $(t_j - t_i) < (y_{count} - threshold)$  then
23          //Too less time between the time instant of the first character and
24          //the time instant of the count-th character
25           $j \leftarrow j + 1$ 
26
27      else if  $(t_j - t_i) < (y_{count} + threshold)$  then
28          //Time correspondence
29           $count \leftarrow count + 1$ 
30           $j \leftarrow j + 1$ 
31
32      else
33          //Too much time between the time instant of the first character and
34          //the time instant of the count-th character.
35          break
36
37  if  $count = |\text{password}|$  then
38      //Time correspondence found
39      return true

```

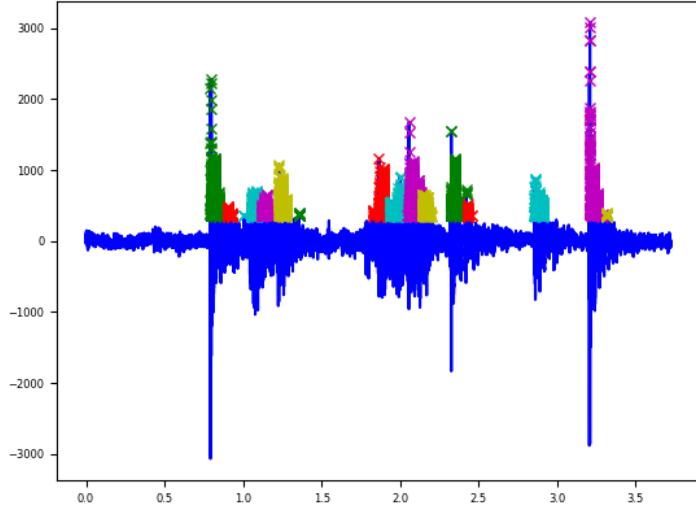


Figure 5.1: Audio during insertion of password `hello35` all sub-windows highlighted.

5.1.2 Character correspondence

The character correspondence also looks for audio peaks during the insertion of the password. The identification of the peaks is performed by using also two threads as mentioned in time correspondence (see **Character 5.1.2**). The thresholding parameters are obtained in the same way, by exploiting the noise peak, but character correspondence also analyses the shape of the audio signal recorded during the insertion of the password.

When a key is pressed by a user on the keyboard, it produces a variation of the audio signal, called *press peak*, for a time window of about 8-10 ms[37]. This signal shape in each window can also be divided in three consecutive and meaningful areas:

- **touch peak**
peak in a window of 2-3 ms, caused by the user's finger touching the key
- **noisy meaningless area**
- **hit peak**
peak in a window of 2-3 ms at the end of 8-10 ms windows, caused by the finger and the key hitting the keyboard supporting plate.

The first goal of character correspondence is to obtain the character related to each key pressed by the user. AcCAPPCHA extract information from the touch peak, that is the most significant, and the hit peak. Following the idea of Asonov and Agrawal, I exploit deep learning to classify each pressed key. In the following sections, there is a detailed explanation of the main phases required and implemented for this classification method:

- Data acquisition
- Extraction of features
- Neural network
- Verification

The first three phases were performed only once and it's not performed by the CAPTCHA. At the contrary the last step is performed at every insertion of the password by AcCAPPCHA. In the first one I record audio files of every key press. From this audio files, I extract information from the peaks in the second phase and then I train a neural network using them in the third step. Hence the verification procedure is based only on: the detection of the audio peaks during the insertion of the password, the extraction of the information from every audio peaks and the classification of them through neural network trained at the previous steps.

5.1.2.1 Data acquisition

The dataset acquisition was performed using two tasks running in parallel: the first one is a key-logger that is used to organize all the recorded audio files in several directories and the second one that records an audio file for each key typed. The communication between the two threads is performed through the private members of the class, in which the thread are created, and the use of the mutual exclusion (mutex). The keylogger continues the acquisition of the audio files even if a special key is pressed (for example F3 button).

The choice of running two tasks in parallel is because the recording session must start before the key press and it must end a little bit later the acquisition of the character. In this way the shape of the signal of the audio peak, related to the key press, are not going to be cut. The data acquisition program can also acquire audio files for a series of key presses but in this case I decided to record one audio file for each key press.

In details the key-logger waits for the insertion of a single key by the user and then reports it to the thread that performs audio recording. This last task

closes the audio stream and stores the audio signal into a *wav* file named with a progressive number. All the audio files are dynamically organized into a set of subfolders of an output directory, each one with the name of the respective typed key.

The recording phase was performed using the built-in Realtek microphone and the keyboard of my MSI GL63 8RD laptop. The names of the subfolders (labels), in which each audio file of a press is inserted, are reported in **Appendix ??**.

Looking at the Table in the Appendix, the keylogger is evident. It maps each key to an ASCII string of upper or lower alphabetic characters because otherwise many keys would be mapped into invalid names of folders (for example, the key '.' is now mapped into the label '*POINT*'). Another observation about the table in **Appendix ??** is that there are two columns of labels: the first one related to the label seen by the key-logger, the second one related to the label manually assigned by me to each key. Sometimes these labels differ for the same entry because of:

- **higher accuracy for spatial distribution of the keys on the keyboard**

for example, '*INSERT*' and '**0_INSERT**' (with Num lock on) would be mapped into '*INSERT*' by the keylogger but they are considered different thanks to my map;

- **improvement of the classification of the keys made by the keylogger**

for example, '*ALT*' label is wrongly mapped into '*SHIFT*' by the keylogger.

- **addition of keys mapped only on the hardware level**

FN is the only key with this problem. The keylogger doesn't detect any key, when *FN* is typed by the user. Hence I had to type *FN*, followed by another key '*a*', so the audio files will be stored in '*a*' subfolder. Then I rename the subfolder '*FN*' and I create a python script to resize all the audio signals inside it and remove the useless second peak, related to '*a*' press.

The last two reasons are very important because they highlight also the power of acoustic side-channel. If an attacker implements an high-level keylogger exploiting also microphone information and remapping wrong labels, the accuracy of its software can increase very much with respect to a traditional keylogger without audio analysis. In fact the hacker could collect a dataset of recordings of pressed keys on the same type of the victim's keyboard and

then it could train a Neural Network to use with malicious purposes. However I record 200 audio signals for each key of my keyboard obtaining a dataset of 20400 audio files. I performed also Data Augmentation on them trying to improve the accuracy of the prediction for the neural network, using two approaches:

- **Time-shift**

from each audio signal I created four new audio signals obtained by applying a time-shift respectively of 0.5, 1.0, 1.5 and 2.0 seconds.

- **Introduction of Gaussian noise**

from each audio signal I created four new audio signals by adding a sequence of random samples, taken from a Gaussian distribution with the standard deviation equal to 150 and the mean equal 0.

Using these approaches I obtained a training set of 1800 audio signals for key, composed respectively by the following datasets:

- 200 audio signals manually recorded by me
- 800 audio signals obtained by time-shift technique
- 800 audio signals from introduction of Gaussian noise

The accuracy of the Neural Network trained on audio signals of both first and second datasets is higher than the one related to the Network trained on first dataset only. The efficiency of the Neural Network trained on first and third dataset is worst than the one related to the network trained on only the first dataset.

I notice that the reason of the previous observation is that the third dataset has many signals, related to different keys, with FFT coefficients similar one to the other. Hence I used only the network trained on the first dataset and both on the first at the second dataset as prediction model. Considering that my keyboard is composed by 102 keys, I used respectively a dataset of 20400 and 102000 audio files to train the neural network.

5.1.2.2 Classification

During the classification phase, I decided how to extract the features from each audio peak, that will be used as input of a neural network. Then I use the features extracted from the audio files in the dataset to train my neural network. used to three different approaches, the first two

I tried three different approaches for the extraction of the features and for each them, the feature for each key is respectively composed by:

- FFT coefficients of the touch peak
- FFT coefficients of the hit peak and the touch peak
- Features obtained from the hit peak and the touch peak using a deep learning pre-trained model

The first two approaches come from the idea in the work of Asonov and Agrawal and the last one was based on the modern sound classification techniques. In the first two cases, the coefficients are extracted from a window of 3 ms around the peaks and then they are normalized in floating point values in range [0, 1] (see **Figure 5.2**).

In the third case, the touch peak and the hit peak samples, taken by 3ms windows, are concatenated and then the spectrogram is computed over these samples (see **Figure 5.3**). From the spectrogram, I extract a feature composed by 512 values through the use of the VGG16 convolutional neural network.

The network was pre-trained on the images in the ImageNet database and I used the intermediate results of neurons, before the last fully connected layers, as feature. The reason of this approach is that a pre-trained network already extracts very well features for good classification of a many labels and it can extract features better than a Convolutional Neural Network for image classification created from scratch.

The features obtained from the audio files are divide into three new dataset: the training set and the test set.

After the extraction of the features with the previous methods, I created a deep neural network from scratch. Its number of input neurons is equal to the size of a feature and the number of output neurons is equal to the number of keys to be mapped. I tried to insert several hidden layers composed by different number of neurons but after changing many parameters, I decided to use only one hidden layer 1024 neurons if the feature was obtained using the spectrograms and 100 otherwise. This choice comes from the best trade-off between the accuracy of the network on the test set and the prediction of labels for new audio peaks obtained later during the insertion of the password.

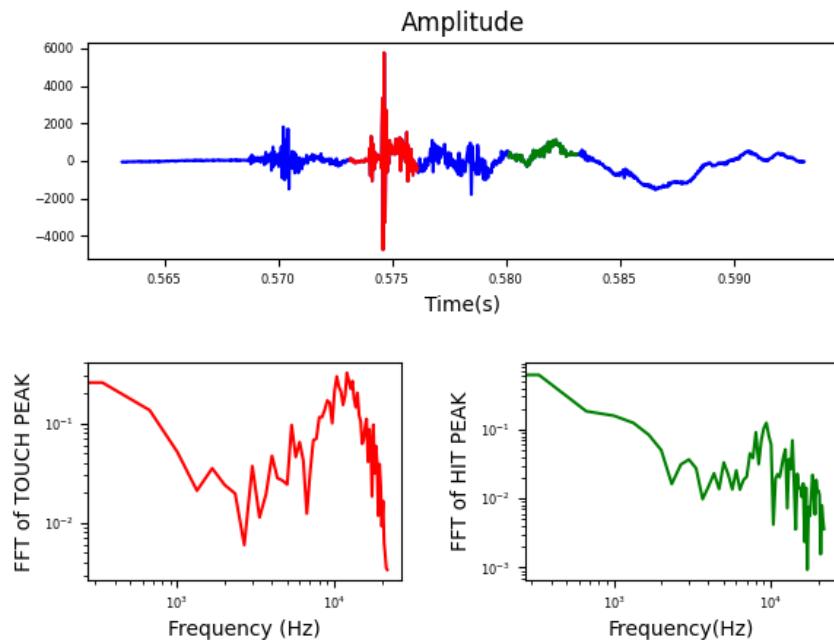


Figure 5.2: Example of normalized FFT computation of the touch peak and the hit peak for an audio file of key 'θ'.

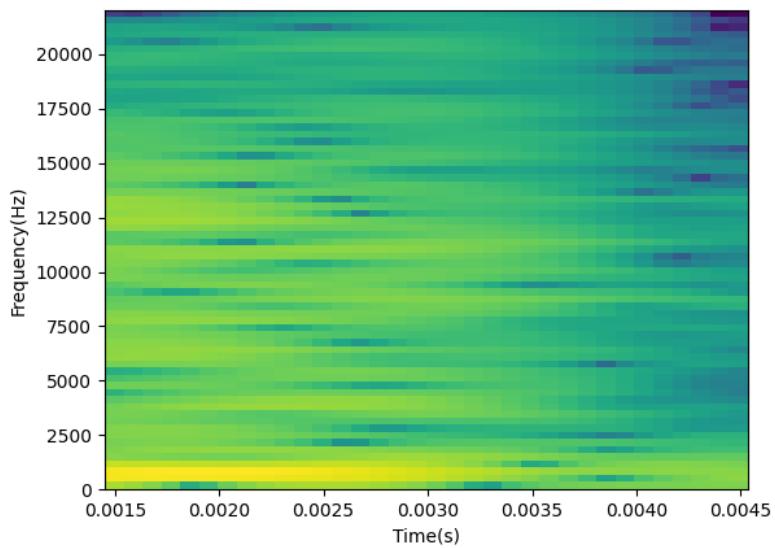


Figure 5.3: Example of spectrogram for an audio file of key 'θ'.

5.1.2.3 Verification

The verification phase is performed by AcCAPPCHA. The audio signal, taken during the insertion of the password, is analysed and then the verification can be based on:

- every audio peak obtained using threshold obtained from the noise evaluation
- audio peaks related to the time instants of to the time correspondence

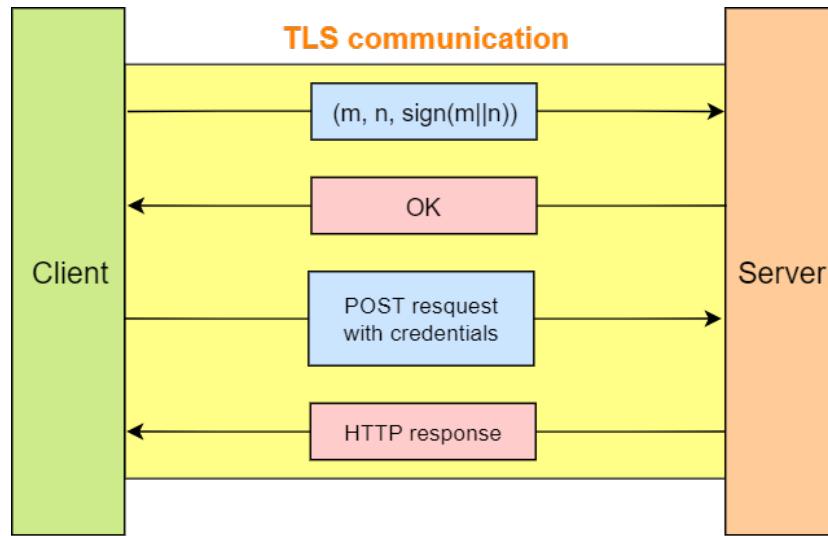
After the selection of the previous peaks, AcCAPPCHA selects the touch and the hit peaks and it computes the features with the approach previously selected by the user from the ones reported in [Section ??](#).

I perform the prediction using the Neural Network. I collect the most probable predicted keys and I repeat the procedure for every window initially computed on the audio. This method is very weak because after this phase, the algorithm tries to find an ordered sequence of characters, one from each window, that corresponds with the password inserted by the user. If this exists, AcCAPPCHA declares that the user was a human, otherwise a bot. The main problem of this approach is that there is no correspondence in time between a character belonging to the final sequence and the moment in which the same character was inserted physically by the user. In fact there can be false positives caused by the prediction from peak that are not related to the absolute maximum one. In other terms, in the set of the maxima of all the windows there can be someone that is not related to the touch peak but to a local maximum.

The second approach solves the previous problem because the windows, where the maxima are looked for, are obtained by the correspondence time approach. In this case, AcCAPPCHA verification becomes more accurate in theory even if in practice the deep learning technique is not very efficient in prediction for a single key.

5.2 Communication between client and server

In the second phase, the username and the password of the user will be signed through ECDSA and sent by client to the authentication service if and only if the insertion was performed by a human. The algorithm that performs the evaluation of user activity (see [Section 5.1](#)) is performed at client side but the value returned by it is evaluated at server-side. The response of the evaluation of user activity concatenated with a nonce and then signed through ECDSA, is sent to the server (see [Section 4.2](#)). The



use of the nonce, unique and random generated sequence, is very important to guarantee that no reply attacks would be performed. In fact the server, after the reception of a message from the client, the server can check if the client has already sent the same nonce before and in this case it declines the message of the client. In this way, any attacker can't reuse a message that previously establishes a client was human. This type of procedure can be also useful to sign HTTP data, for example data sent using POST request (as insertion of password during authentication phase). In the testing phase I performed, I designed and implemented also a simplified version of the communication between the client and the server for an authentication service.

The application was tested on local network and the actions performed by involved parties are described in details in the following sections (see Figure).

5.2.1 Client

The client performs the authentication following these steps:

1. It establishes a connection with the server over TLS layer to increase the security strength of the communication between the parties;
2. It sends the message $(m, n, \text{sign}(m||n))$ where:
 - m is the string with the response of AcCAPPCHA algorithm on client side ('True' if Human, False if bot)
 - n is the nonce

- $sign(m||n)$ is the ECDSA signature of the concatenation $m||n$ of the response and the nonce. For ECDSA was used SHA256 algorithm.

From a practical point of view, I format the message in the following way:

```
m CRLF n sign(m|n)
```

According to basic rules in grammar of **HTTP/1.1** (see Section 2.2 of RFC 2616), *CR* is the carriage return ('\r') and *LF* is the line feed ('\n'). The spaces in the message aren't considered. In this way, I can separate easily *m* looking at '\r \n'. The nonce has a fixed length of 16 bytes.

3. The client waits for response of the server with format:

```
response CRLF
```

If the answer is equal '*OK \ r \ n*', AcCAPPCHA will go on with the authentication step, otherwise the client-side application performs again the verification, asking again to user to insert the password. The maximum number of trials for a particular user is 3.

4. If everything goes well in the previous step, The client sends the credentials (username, password) to the server through a HTTP POST request to '/cgi-bin/auth' resource. The name of folder '/cgi-bin' comes from the standard name of the folder with functions and *auth* is the name of the function that server will call. This naming approach was used very much in the past to separate functions code from pure HTML code. The password is not sent directly but it's hashed before using SHA512. The POST request used by the client has the following format:

```
POST /cgi-bin/auth HTTP/1.1 \r \n
Host: SP foo.example CRLF
Content-Type: SP application/x-www-form-urlencoded CRLF
Content-Length: SP SIZE CRLF CRLF
user = USERNAME & pwd = HASHPWD
```

where everything follows as before the grammar in RFC 2616). In fact also **SP** represents the space character as in the documentation. **SIZE**, **USERNAME** and **HASHEDPWD** are replaced respectively with the size of the HTTP body, the username of the client and its password hashed with SHA512.

5. The client waits for the HTTP response of the server, containing HTML code as body. Then the client saves the code on the file system and opens the default web browser only to show. The HTML code is intended to show 3 possible scenarios: the user was correctly logged in, the user inserted wrong password, the username wasn't already stored on the server database.

5.2.2 Server

The client performs the authentication following these steps:

1. It establishes a connection with the client, after his request, over TLS layer to increase the security strength of the communication between the parties;
2. It receives the message $(m, n, sign(m||n))$ and check the integrity of the message. To do it, the server decrypts the ECDSA signature using the client's ECDSA public key and compares the result with $m//n$.
3. If compared messages were the same, the server checks if the nonce was already used by the same client. If so, the server thinks that there was an attacker that is performing a replay attack. If the nonce wasn't already used by the client, it will be stored in a dictionary to monitor clients activity. Each entry of the dictionary is composed by:

- **Key: IP address**

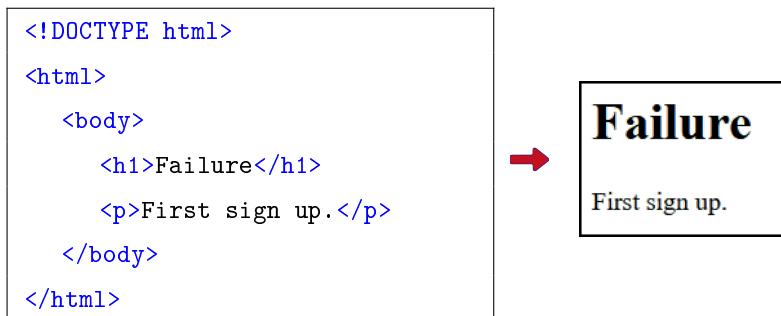
It is the IP address of the client and it is a simplification of the information that identifies a client. For example the client could be associated also to port number used to make the request, the Operating System on which the AcCAPPCHA was running on client-side or other useful parameters.

- **Value: list of nonces**

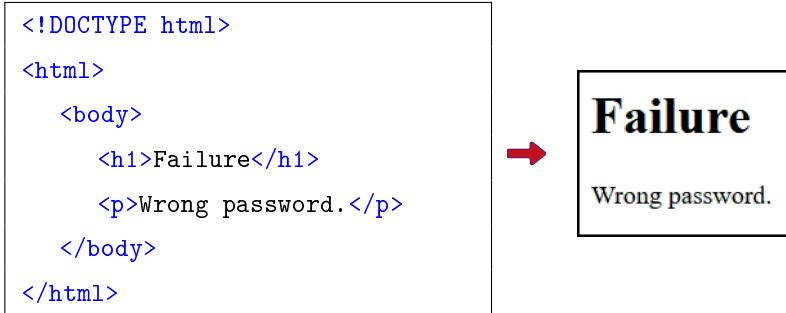
Every time a client performs a new verification request on the server, the nonce is added to the list related to its IP address in the dictionary.

4. If the nonce was used for the first time by the client, the server checks the value of the response received by the client. If the response is 'True', the server replies 'OK \ r \ n' otherwise 'NO \ r \ n'. If some error occurs it sends 'ERROR \ r \ n' to the client. In the last two cases, the server terminates.

5. If the server doesn't terminate, it waits for the POST request from the client and analyses it to perform authentication service. The server replies to client with several status codes:
 - **501** (Not implemented)
If the request is not a POST (e.g. GET)
 - **400** (Bad Request)
If the number of the parameters in the POST body is different from 2 (username and password).
 - **200** (OK)
If the number of the parameters in the POST body is equal to 2, the server will reply with an HTTP response with a body content depending on several cases:
 - **User not in the database**
the server sends the following HTML code if the specified username isn't already stored in the database.

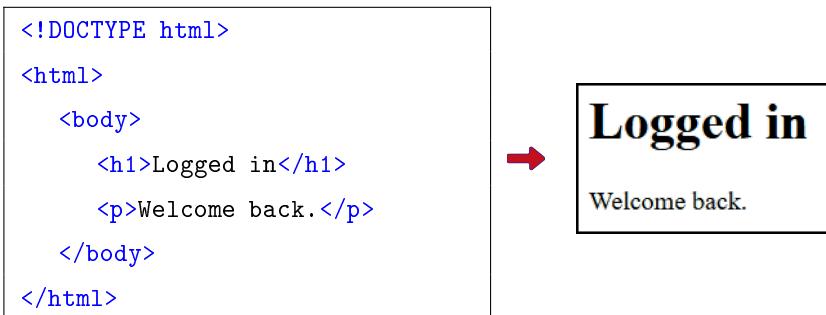


- **Wrong password**
the server sends the following HTML code if the hashed password received from the client isn't the same with respect to the one stored in database for the specified username.



- **User logged in**

the server sends the following HTML code if the specified user-name exists in the database and its hashed password stored in the database is the same of the one received through POST request.



5.2.3 Database

The database, created to simulate the search of a username by the server, was made using PostgreSQL. I could store all the information in a simple text file or a csv file but I decide to use this approach to be more flexible to future integration to more complex database.

The database is composed by only a table `CloudUser` to store information about user identity usually stored during the sign up phase. The creation of the database was performed through the following instructions:

— Database Creation
`CREATE DATABASE cloudservice OWNER POSTGRES ENCODING = 'UTF8';`

— Connect to `cloudservice` db to create data for its '`public`' schema
`\c cloudservice`

— Create new domains

```

— Correct password format
CREATE DOMAIN pwd AS char(128)
  CONSTRAINT properpassword CHECK (((VALUE)::text ~* '[a-f0-9]' :: text));

— Correct password format
CREATE DOMAIN userformat AS char(128)
  CONSTRAINT properpassword CHECK (((VALUE)::text ~* '[A-Za-z0-9]' :: text));

— Correct mail format
CREATE DOMAIN mail AS character varying(254)
  CONSTRAINT propermail CHECK (((VALUE)::text ~* '[A-Za-z0-9._%-]+@[A-Za-z0-9._%]+[$' :: text)));

— Create new data type
CREATE TYPE gendertype AS ENUM (
  'Male',
  'Female'
);

— Create tables
CREATE TABLE CloudUser (
  Name VARCHAR NOT NULL,
  Surname VARCHAR NOT NULL,
  Username userformat NOT NULL,
  Email mail NOT NULL,
  Sex gendertype,
  Password pwd NOT NULL,
  PRIMARY KEY (Username)
);

```

I created several domain to manage format of some information of the user. For example the password is a string of 128 characters because each password is hashed using SHA512 and then formatting the result as a string of hexadecimal values. Then I populated the dataset with some entries, related to fake user, only for testing purpose. An example of inserted users is the following one:

```

INSERT INTO CloudUser (Name,
                      Surname,
                      Username,
                      Email,
                      Sex,
                      Password)

VALUES ('Raffaele',
        'Papini',
        'raffaele.papini',
        'raffaele.papini@outlook.it',
        'Male',
        '12345678901234567890123456789012');

```

```
'Di Nardo Di Maio',
'RaffaDNDM',
'example1@gmail.com',
'Male',
HASHPASSWORD);
```

In practice, HASHPASSWORD is replaced by the string of 128 characters corresponding to the hashed password in hexadecimal format.

5.2.4 Encryption Keys

The creation of the TLS socket for the communication between the client and the server is done by using keys and certificates created thanks to the following bash instructions:

```
openssl req -new -x509 -days 365 -nodes -out client.pem
            -keyout client.key
```

```
openssl req -new -x509 -days 365 -nodes -out server.pem
            -keyout server.key
```

OpenSSL is a open-source implementation of TLS/SSL protocols and, thanks to the option `-x509`, you can display certificates and also access to many signing protocols. In particular, in the previous bash instructions, a X.509 Certificate Signing Request (CSR) is generated and signed for both the parties.

Thanks to `-nodes` the private key is created and not encrypted. The certificates are stored respectively in `server.pem` for the server side and `client.pem` for the client and they are valid for 365 days. The private keys are stored thanks to `-keyout` option in the `client.pem` and `server.key`.

The keys, used in ECDSA signing and verification, were created as follow from Python language instead of using a bash tool:

```
from ecdsa import *
from hashlib import sha256

PRIVATE_KEY = SigningKey.generate(curve=SECP256k1,
                                  hashfunc=sha256)

with open('ecdsa.key', 'w') as private_pem:
    private_pem.write(PRIVATE_KEY.to_pem().decode())

PUBLIC_KEY = PRIVATE_KEY.get_verifying_key()
```

```
with open('ecdsa.pem', 'w') as public_pem:  
    public_pem.write(PUBLIC_KEY.to_pem().decode())
```

The *ecdsa* module gives access to the management of operations performed by signing and verification phase. In this case the private key, used to sign a message from the client, was computed on the curve SECP256k1 usually used in Bitcoin applications.

Chapter 6

Experimental results

All the tests were performed on a MSI GL63 8RD laptop, using Windows 10 Home version 1909 as Operating System.

6.1 Human detection

The application response was tested over all the three possible scenarios (see [Section 5.1](#)):

- Time correspondence
- Character correspondence
- Both time and character correspondence

The first method is the most efficient and usually at first or second trial already finds a human activity. Sometimes the correspondence isn't found at first trial because of background noise.

The last two approaches based on deep learning require the reduction of the number of labels to obtain good results. In this way I limit the possible characters that user could use and the password should be composed of lower case alphabetic characters, accented vowels or numbers (see [Figure 6.1](#)) but considering again the 10 high probable labels predicted for each peak.

In any case the last method that uses both the time and the character correspondence speed up the search of the correct peak related to the insertion and reduces the possibility of having a false positive. However there are many false negatives given by the similar sound produced by several keys opposite to D. Asonov, R. Agrawal's conclusions[[37](#)].

The reasons of their presence could be:

- different parts of my keyboard plate produce similar sounds
- the built-in microphone of my laptop is more affected by the noise than an external microphone
- the movements of the user's hands influence the noise during the password insertion
- some keys are more worn out than other. The worst deep learning method is the one that uses the spectrograms and a pre-trained deep learning model to extract the features of a press peak. In **Table ??** However during the insertion of the password, the user can obviously type backslash key but then '\b' character isn't considered in the final verification.



Figure 6.1: Keys to be used in the password (highlighted in green).

Feature type	<i>bye5</i>	<i>hello35</i>
	2 (3/4)	1
FFT of touch peak	2 (3)	1
	1	1
	2 (3)/4	2 (6/7)
FFT of touch and hit peaks	2 (3/4)	1
	3 (3/4)	3 (5.5/7)
	NO (3)	NO (5.33/7)
Spectrogram of touch and hit peaks	NO (2.67/4)	NO (6/7)
	NO (2.67/4)	NO (4.67/7)

6.2 Bot detection

I've tested both human and bot activities analysing the response generated by AcCAPPCHA. The bot activity was emulated using several approaches and using the username '**RaffaDNDM**' and its password '**hello35**:

- **Python program with popen communication**

this approach opens a subprocess and a communication through pipes with stdin and stdout streams (see). Assuming the hacker obtained the credentials in some way, the bot communicates the username and the password to running AcCAPPCHA with only time correspondence option selected. In practice the strength of AcCAPPCHA against this attack is very high because the insertion of the password is managed through `getwch()` call in the Windows Operating system.

This function belongs to `msvcrt` module and takes one character at the time. This module guarantees also that the console I/O routines are not compatible with stream I/O or low-level I/O library routines. In the Windows operating systems, the output from these functions is always directed to the console and cannot be redirected through any king of pipes.

The program easily accepts the username, because AcCAPPCHA acquires it using standard `input()` function. Then AcCAPPCHA waits for characters of password, until ENTER key is pressed, but `popen` doesn't have access to the stream analysed by `getwch()`. I didn't iterate the insertion of the password for the maximum number of possible trials because even at the first insertion, AcCAPPCHA doesn't see the insertion of the bot.

```
from subprocess import Popen
import sys
import msvcrt
from time import sleep

def popen_bot(username, password):
    #Subprocess that redirects pipes
    process = Popen('python3 AcCAPPCHA.py -t -plot',
                    shell=True,
                    stdin=subprocess.PIPE,
                    stdout=subprocess.PIPE,
                    stderr=subprocess.STDOUT)

    #Wait until username could be inserted
    sleep(4)
    #Write username and password
    credentials = username.encode() + b'\r\n' + \
```

```

        password . encode () + b '\r\n'
output = process . communicate (credentials) [0]

print (output . decode ())

```

Listing 6.1: Bot using popen.

- **Python program with pyinput module**

Using this module, I pretended to be a bot and access directly the console, bypassing the stream limits of **msvcrt** character acquisition. To emulate the user the bot program should start the execution run and immediately after the hacker must open the working terminal with AcCAPPCHA running.

This scenario isn't very feasible because requires management of terminal windows but was useful to establish if the insertion of the password by malicious software is correctly classified as a bot activity. Each character of the password is inserted emulating press and release of the corresponding key of the keyboard.

```

from pyinput . keyboard import Key , Controller
from time import sleep

def input_bot (username , password):
    #Object for control of keyboard events
    keyboard = Controller()

    def press_release (char):
        keyboard . press (char)
        keyboard . release (char)

    #Wait that username could be inserted
    sleep (4)

    #username insertion
    for x in username:
        press_release (x)

    press_release (Key . enter)

    #Trials for password insertion
    count = 0
    while (count < 3):
        sleep (5)

        for x in password:
            press_release (x)

```

```

press_release(Key.enter)
count += 1

```

Listing 6.2: Bot using pyinput module.

- **Remote control of the PC**

the last test was performed by using the program Team Viewer and accessing directly the terminal. For this reason, this type of attack isn't feasible in practice as the previous mentioned attack with Python bot.

In the last two approaches the audio recorded by AcCAPPCHA are very similar and highlights the two most probable situations:

- **The noise during noise evaluation is very high**

If this happens, no audio peaks can be found (see **Figure 6.2**).

- **The noise during noise evaluation is normal or very low**

If this happens, some audio peaks can be found if there is some noise during password insertion. However they don't have time correspondence with time instants stored during the password insertion (see **Figure 6.3**). The only case, in which a bot can authenticate it self, is when there is a sequence of audio peaks caused by the noise and the time between them is the same of the stored ones. However this event isn't very probable because it is hard that there would be too high noise only during the insertion of the password and not during the noise evaluation. For example, if the attacker would analyse the background noise finding a sequence of peaks and defining a model for amount of time between each couple of them, he couldn't use information from it. In fact it could insert a character of the password after an amount time previously modelled but in practice, the peaks of background noise would be discarded through noise evaluation.

6.3 Usability

The verification with deep learning techniques would be more strong if the predictions would be more accurate. A problem of this methods is that the neural network must be trained using audio file of keys of a specific keyboard. This method limits the usability of the resulting weights for all the neurons to the specific keyboard previously used.

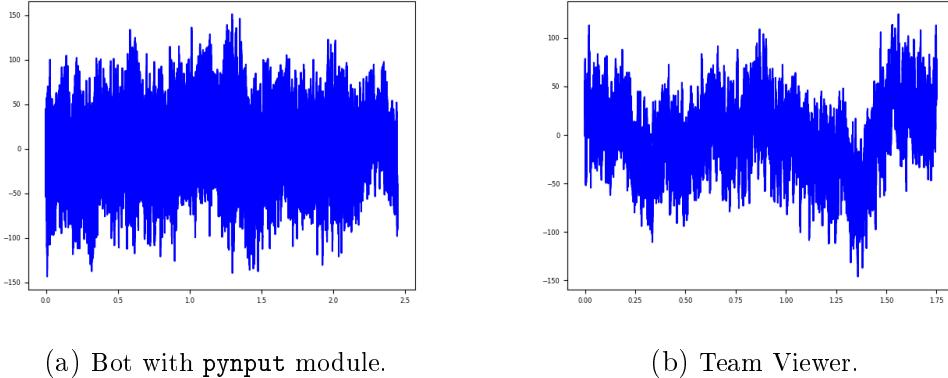


Figure 6.2: Plot of audio during the password insertion with high noise during noise evaluation.

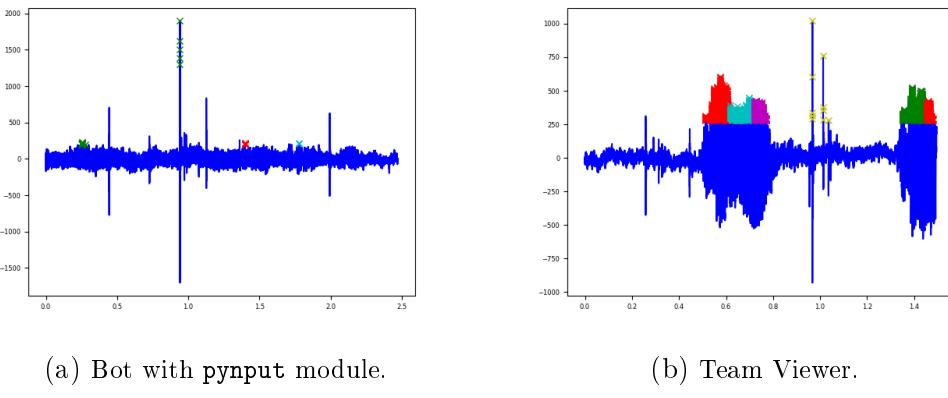


Figure 6.3: Plot of audio during the password insertion with low noise during noise evaluation.

To use the character correspondence on every possible keyboard, each producer of keyboards should record audio produced by pressed keys and share the results to increase the efficiency of AcCAPPCHA. This action should also bring to a malicious use of the data (e.g. an attacker could use them to produce an advanced key-logger).

I used also other input devices to test the usability of the best method (time correspondence). At first I tested AcCAPPCHA using an external wireless keyboard (Logitech K480) by positioning it in front of my laptop. I tested AcCAPPCHA using touch keyboard and one the following hardware devices:

- **Mouse**

this device guarantees to perform the time correspondence exploiting

the sound of each mouse click needed to select a key.

- **Touchpad**

this device also guarantees to perform the time correspondence exploiting the sound of each click the touchpad needed to select a key. The only condition, required by AcCAPPCHA to work correctly, is that the user clicks the physical left button at the bottom side of the touchpad.

The results of verification using these three devices were the same of the ones obtained using the keyboard of the laptop.

6.4 Security analysis

Following the analysis performed for Invisible CAPPCHA[35], I'm going to discuss the strength of AcCAPPCHA against the following attacks:

- **Replay attack**

The message is still concatenated with a nonce and then they are signed to guarantee a client would use a nonce only one. The server still prevents this type of attacks by refusing the second message from a client with a nonce already used by him.

- **Reverse engineering attack**

The code working on the client must be kept secure by the File System. Even if the attacker could reverse code on server, the signature of the message and the TLS communication guarantee that this attack cannot be performed.

- **Human-solver relay attack**

AcCAPPCHA is strong there aren't additional tasks to be performed as in classical CAPTCHAs. Hence no challenges can be sent to remote human solver.

- **Brute force and password replay attacks**

at the contrary of Invisible CAPPCHA, a single call to AcCAPPCHA performs the verification of human identity for at most 3 times:

- **if a bot is detected in one of the 3 attempts**

the program asks again the password to the user and reduces the number of remaining trials

- **Otherwise**

the password inserted by the user is compared with the one in the

database. In any case, the client program terminates receiving an HTML page with the result of the login phase.

The server doesn't know if AcCAPPCHA was called more than once by the client. If the server wants to allow no more than a fixed amount of attempts, it could count the number of POST requests received during a fixed amount of time by each client. If the number of requests reaches the maximum number of designed attempts, the server adds the corresponding client to a black list and then communicates it to the client. Hence the client-side of AcCAPPCHA, receiving the message from server, will refuse next attempts of verification (insertion of password).

- **Denial Of Service (DOS)**

For the same reason previously explained, the number of passwords received by the server is 1. To prevent a DOS attack, the server could use the same approach with black list designed to prevent the previous attack. In fact the server won't be overloaded, because the client will refuse next attempts.

The choice of the HTTP POST method is because it's more secure to exchange sensible information. In fact, an HTTP GET method is easily associated to an URL that could be easily stored in the Browser History.

File System -> keys (C+S), block.txt+spectrograms (C) If an attacker has remote control of the pc on which client-side AcCAPPCHA would run, the user could understand on Windows 10 that he's under attack by looking the microphone icon on the taskbar. In fact, I suggest every user to enable this icon in **Settings > Personalisation > Taskbar > Turn system icons on or off** even if an attacker could disable it. Every time that an application accesses the microphone resource, the user could see the icon on the taskbar if it was enabled on the settings.

Chapter 7

Future work

In the future, AcCAPPCHA could be tested on several Operating Systems to make sure that it works correctly on all of them. The application could still be improved by collecting more audio files and increasing the accuracy of the classification approach based on the spectrograms. We can add the management of a blacklist on server side of the AcCAPPCHA to prevent Brute Force attacks and DOS attacks, as explained in [Section 6.4](#).

The most important work that could be added is the implementation of time correspondence approach on smartphones. We can exploit the analysis of audio signals recorded by the microphones of the mobile phones. In fact, many smartphones are now equipped with two microphones and these can be used to increase the accuracy during the research of the audio peaks.

Another task that could be added to the mobile version of AcCAPPCHA is the character correspondence. It could exploit the shape of the waves but also the time difference between the two audio peaks, related to the same typed character, and recorded by the built-in microphones[[38](#)].

Moreover the voice assistants, like Amazon Echo and Google Home, are becoming increasingly widespread and they are equipped with several microphones. These devices record human activity continuously and only if the user says a specific keyword, they turn on speakers and reply to user. Hence, the audio signals could be exploited to develop new side-channel attacks and analyse input on physical or virtual keyboards[[39](#)]. Hence the collected information could be exploit to develop also new version of AcCAPPCHA with these devices.

VALIDATION SET FOR IMPROVEMENT

Appendix A

Key Map



Figure A.1: Layout of the keyboard of my MSI GL63 8RD laptop.

Key	Key-logger label	Final label
	0	0
	<ul style="list-style-type: none">• INSERT (Num lock on)• None (Num lock off)	0_INSERT
	1	1
	<ul style="list-style-type: none">• END (Num lock on)• None (Num lock off)	1_END
	2	2
	<ul style="list-style-type: none">• DOWN (Num lock on)• None (Num lock off)	2_DOWN

	3	3
	<ul style="list-style-type: none"> • PAGE_DOWN (Num lock on) • None (Num lock off) 	3_PAGE_DOWN
	4	4
	<ul style="list-style-type: none"> • LEFT (Num lock on) • None (Num lock off) 	4_LEFT
	5	5
	<ul style="list-style-type: none"> • None (Num lock on) • None (Num lock off) 	5_NUM
	6	6
	<ul style="list-style-type: none"> • RIGHT (Num lock on) • None (Num lock off) 	6_RIGHT
	7	7
	<ul style="list-style-type: none"> • HOME (Num lock on) • None (Num lock off) 	7_HOME
	8	8
	<ul style="list-style-type: none"> • UP (Num lock on) • None (Num lock off) 	8_UP
	9	9
	<ul style="list-style-type: none"> • PAGE_UP (Num lock on) • None (Num lock off) 	9_PAGE_UP
	a	a
	SHIFT	ALT
	CTRL	ALT_GR

	APOSTROPHE	APOSTROPHE
	b	b
	BACKSLASH	BACKSLASH
	BACKSPACE	BACKSPACE
	c	c
	CAPS_LOCK	CAPS_LOCK
	COMMA	COMMA
	CTRL	CTRL
	CTRL_R	CTRL_R
	d	d
	DELETE	DELETE
	DOWN	DOWN
	e	e
	ENTER	ENTER
	ENTER	ENTER_R
	ESC	ESC

	f	f
	F1	F1
	F2	F2
	F3	F3
	F4	F4
	F5	F5
	F6	F6
	F7	F7
	F8	F8
	F9	F9
	F10	F10
	F11	F11
	F12	F12
		FN
	g	g
	h	h
	i	i
	INSERT	INSERT
	j	j
	k	k

	l	l
	LEFT	LEFT
	LOWER	LOWER
	m	m
	MINUS	MINUS
	MINUS	MINUS_R
	n	n
	NUM_LOCK	NUM_LOCK
	o	o
	p	p
	PAGE_DOWN	PAGE_DOWN
	PAGE_UP	PAGE_UP
	PAUSE	PAUSE
	PLUS	PLUS
	PLUS	PLUS_R
	POINT	POINT
	<ul style="list-style-type: none"> • DELETE (Num lock on) • None (Num lock off) 	POINT_DELETE

	PRINT_SCREEN	PRINT_SCREEN
	q	q
	r	r
	RIGHT	RIGHT
	s	s
	SCROLL_LOCK	SCROLL_LOCK
	SHIFT	SHIFT
	SHIFT	SHIFT_R
	MINUS	SLASH
	SPACE	SPACE
	STAR	STAR
	t	t
	TAB	TAB
	u	u
	UP	UP
	v	v
	w	w

	SHIFT	WINDOWS
	x	x
	y	y
	z	z
	à	à
	è	è
	ì	ì
	ò	ò
	ù	ù

Table A.1: Map of pressed key performed by key-logger and then manually modified in final label.

Appendix B

Program

Bibliography

- [1] 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, 2013.
- [2] Darko Brodić, Alessia Amelio, Radmila Janković, "Exploring the influence of CAPTCHA types to the users response time by statistical analysis" in *Multimedia Tools and Applications*, vol. 77, pp. 12293–12329, 2017.
- [3] Walid Khalifa Abdullah Hasan, "A Survey of Current Research on Captcha" in *International Journal of Computer Science Education in Schools (IJCSES)*, vol. 7, pp. 141–157, 2016.
- [4] Ruti Gafni, Idan Nagar, "CAPTCHA – Security affecting user experience" in *Issues in Informing Science and Information Technology*, vol. 13, pp. 63-77, 2016.
- [5] G. Sauer, J. Holman, J. Lazar, H. Hochheiser, and J. Feng, "Accessible privacy and security: A universally usable human-interaction proof tool" in *Univers. Access Inf. Soc.*, vol. 9, no. 3, p. 239–248, Aug. 2010.
- [6] Jennifer Tam, Jiri Simsa, David Huggins-Daines, Luis von Ahn, Manuel Blum, "Improving Audio CAPTCHAs" in *Symposium On Usable Privacy and Security (SOUPS)*, 2008.
- [7] William Aiken, Hyoungshick Kim, "POSTER: DeepCRACK: Using Deep Learning to Automatically CRack Audio CAPTCHAs" in *Proceedings of the 2018 on Asia Conference on Computer and Communications Security (ASIACCS)*, 2018.
- [8] Jennifer Tam, Jiri Simsa, Sean Hyde, Luis von Ahn, "Breaking audio CAPTCHAs" in *Advances in Neural Information Processing Systems*, pp. 1625–1632, 2009.

- [9] Manar Mohamed, Song Gao, Nitesh Saxena, Chengcui Zhang, "Dynamic cognitive game captcha usability and detection of streaming-based farming" in *The Workshop on Usable Security (USEC), co-located with NDSS, 2014.*
- [10] Ved Prakash Singh, Preet Pal, "Survey of Different Types of CAPTCHA" in *International Journal of Computer Science and Information Technologies (IJCSIT), vol. 5(2), 2014.*
- [11] Goswami G, Powell BM, Vatsa M, Singh R, Noore A, "FaceD-CAPTCHA: face detection based color image CAPTCHA" in *Future Generation Computer Systems, vol. 31(2), pp. 59–69, 2014.*
- [12] Wen-yao Lu, Ming Yang, "Face Detection Based on Viola-Jones Algorithm Applying Composite Features" in *International Conference on Robots & Intelligent System (ICRIS), pp. 82-85, 2019.*
- [13] Brian M. Powell, Abhishek Kumar, Jatin Thapar, Gaurav Goswami, Mayank Vatsa, Richa Singh, Afzel Noore, "A multibiometrics-based CAPTCHA for improved online security" in *IEEE 8th International Conference on Biometrics Theory, Applications and Systems, 2016.*
- [14] Carlos Javier Hernandez-Castro, Arturo Ribagorda, "Pitfalls in CAPTCHA design and implementation: The Math CAPTCHA, a case study" in *Computers & Security, vol. 29(1), pp. 141-157, 2010.*
- [15] Luke Wroblewski, "A Sliding Alternative to CAPTCHA?", 2010.
- [16] Silky Azad, Kiran Jain,, "Captcha: Attacks and weaknesses against OCR technology" in *Global Journal of Computer Science and Technology, 2013.*
- [17] E. Bursztein, M. Martin, J. Mitchell, "Text-based CAPTCHA strengths and weaknesses" in *Proc. 18th ACM Conference on Computer and Communications Security (CCS), pp. 125–138, 2011.*
- [18] M. Shirali-Shahreza and S. Shirali-Shahreza, "Motion CAPTCHA" in *2008 Conference on Human System Interactions, Krakow, pp. 1042-1044, 2008.*
- [19] Kameswara Rao Kavya Sri, Gnana Sai, "A Novel Video CAPTCHA Technique to Prevent BOT Attacks" in *International Conference on Computational Modeling and Security (CMS 2016), Procedia Computer Science, vol. 85, pp. 236–240, 2016.*

- [20] K. A. Kluever, R. Zanibbi, "Balancing usability and security in a video captcha" in *Proceedings of the 5th Symposium on Usable Privacy and Security, 2009.*
- [21] Omar Ahmed Hedaia, Ahmed Shawish, Hana Houssein, Hala Zayed, "Bio-CAPTCHA Voice-Based Authentication Technique for Better Security and Usability in Cloud Computing" in *International Journal of Service Science Management Engineering and Technology, vol. 11(2), pp. 59-79, 2020.*
- [22] Erkam Uzun, Simon Chung, "rtCaptcha: A Real-Time Captcha Based Liveness Detection System" in *The Network and Distributed System Security Symposium (NDSS), Georgia Institute of Technology, 2018.*
- [23] Vinay Shet, "Are you a robot? Introducing "No CAPTCHA re-CAPTCHA"" , 2014.
- [24] TehnoBlog, "Google no Captcha + INVISIBLE reCaptcha–First Experience Results Review", 2019.
- [25] Suphanee Sivakorn, Jason Polakis, Angelos D. Keromytis, "I'm not a human: Breaking the Google reCAPTCHA" in *Black Hat, pp. 1–12, 2016.*
- [26] Meriem Guerar, Alessio Merlo, Mauro Migliardi, "Completely automated public physical test to tell computers and humans apart: A usability study on mobile devices" in *Future Generation Computer Systems, vol. 82, pp. 617–630, 2018.*
- [27] Narges Roshanbin, James Miller, "A survey and analysis of current CAPTCHA approaches" in *Journal of Web Engineering, vol. 12, issue 1–2, pp. 1–40, 2013.*
- [28] R. Spreitzer, V. Moonsamy, T. Korak and S. Mangard, "Systematic Classification of Side-Channel Attacks: A Case Study for Mobile Devices" in *IEEE Communications Surveys & Tutorials, vol. 20, no. 1, pp. 465–488, 2018.*
- [29] A. Nahapetian, "Side-channel attacks on mobile and wearable systems" in *13th IEEE Annual Consumer Communications & Networking Conference (CCNC), pp. 243-247, 2016.*
- [30] Paul Kocher, Joshua Jaffe, Benjamin Jun, "Introduction to Differential Power Analysis and Related Attacks" in *Cryptography Research, 1998.*

- [31] J. W. Bos, C. Hubain, W. Michiels, and P. Teuwen, "Differential Computation Analysis: Hiding Your White-Box Designs is Not Enough" in *Cryptographic Hardware and Embedded Systems – CHES, ser. LNCS, vol. 9813, Springer, pp. 215-236, 2016.*
- [32] S. Skorobogatov, "The Bumpy Road Towards iPhone 5c NAND Mirroring" in *arXiv ePrint Archive, Report 1609.04327, 2016.*
- [33] Shuo Chen, Rui Wang, XiaoFeng Wang, Kehuan Zhang, "Side-Channel Leaks in Web Applications: A Reality Today, a Challenge Tomorrow" in *2010 IEEE Symposium on Security and Privacy, pp. 191-20, 2010.*
- [34] J. Zhang, X. Zheng, Z. Tang, T. Xing, X. Chen, D. Fang, R. Li, X. Gong, and F. Chen, "Privacy Leakage in Mobile Sensing: Your Unlock Passwords Can Be Leaked through Wireless Hotspot Functionality" in *Mobile Information Systems, vol. 2016, pp. 8793025:1–8793025:14, 2016.*
- [35] M. Guerar, A. Merlo, M. Migliardi, F. Palmieri, "Invisible CAPTCHA: A usable mechanism to distinguish between malware and humans on the mobile IoT" in *Computers & Security, vol. 78, pp. 255–266, 2018.*
- [36] J. V. Monaco, "SoK: Keylogging Side Channels" in *2018 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 2018, pp. 211-228.*
- [37] D. Asonov, R. Agrawal, "Keyboard acoustic emanations" in *IEEE Symposium on Security and Privacy. Proceedings. 2004, Berkeley, CA, USA, 2004, pp. 3-11.*
- [38] I. Shumailov, L. Simon, J. Yan, R. Anderson, "Hearing your touch: A new acoustic side channel on smartphones" in *arXiv, 2019.*
- [39] A. Zarandy, I. Shumailov, R. Anderson, "Hey Alexa what did I just type? Decoding smartphone sounds with a voice assistant" in *arXiv, 2020.*

Acknowledgements

Professor Migliardi

I would like to express my very great appreciation to Dr Guerar

I would like to express my gratitude to University of Padova for the study path I performed. The uncertainty about the future and the idea of being far from needed cyber security skills have become a stimulus to improve myself. I learn a lot and I got hooked on the programming, starting from zero level of it, thanks to the professors' professionalism and knowledge. During the last five years, I've changed and now I spend programming all my free time. Thanks to University because professors follow my thirst of knowledge and I grew up, living alone and really becoming an adult.

Thanks to staff,

Thanks to other company,

I would like to express my gratitude to my family that taught me to never give up. In particular thank to my sister that, with her great experience in University course, has been a reference for my study attitude and perseverance.

Thanks to my grandmother Concetta that, with her smiles, has left to me happiness during my darker and difficult periods.

Thanks to my grandmother Maria that, was always worried for my health issues and tries always to give me up using her food.

Thanks to Cristina,

Thanks to Francesca, because even if we had many commitments we have

always found 5 minutes to stay together and to support the other one in his/her choices.

Thanks to Davide,

Thanks to Giuseppe, Aurora, Alessia and Sara,

Thanks to Elia,

Thanks to Lorenzo,