Authentication by Gesture Recognition: a Dynamic Biometric Application

Submitted by

Benoit DUCRAY

for the degree of Doctor of Philosophy

of the

Royal Holloway, University of London

2017

Declaration

I, Benoit DUCRAY, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed . . . Ducray . . . (Benoit DUCRAY)
Date:17/05/2017

Abstract

With the expansion of digital information and the number of people potentially able to access it, there are increasing demands for efficient, secure systems which authenticate users effectively. At the end of the 1960s, IBM defined three authentication factors: knowledge factor, which relates to "something the entity knows"; ownership factor, relating to "something the entity has/possesses"; inherent factor, which can be summarised as "something the entity is or does".

Each of these factors possesses its own limitations: knowledge factors can be forgotten or discovered by a fraudster; ownership factors can be lost, stolen or counterfeited. These nuisances have led to increased use of biometric authentication as a means of increasing security. However, conventional biometrics are static, so if compromised cannot be changed by the user. This has led to interest in techniques that authenticate using changeable multi-factor authentication measures that are influenced by biometrics, rather than being completely reliant on them. Investigating the practicality and security of such techniques provided the motivation for this research.

First we needed to identify these "easily changeable biometrics" and it led us to define a new biometric family, which we called the dynamic biometrics. We then studied the security characteristics of this new family.

Out of the dynamic biometrics, we chose to focus on one of these elements, authentication based on gesture recognition. We conducted several experiments to assess the ability of this authentication technique to authenticate the genuine user and reject any impostors, either if these impostors do not know the gesture or do random movements.

We continued by looking for a secure place to store the gesture recognition's template and run the application. We evaluated the possibility of doing that on a personal limited device, such as a Smart Card.

Then we designed a protocol to use a gesture recognition application which we analysed with respect to several threat vectors.

Acknowledgement

This work would never have come through without the help of many I received along the years.

Firstly, I would like to thank Professor Keith Mayes for the opportunity of joining RHUL's Smart Card Centre team, as well as for the time and support he offered me when supervising my work.

My thoughts also go to Sheila Cobourne, whose time spent to helping me with the English language has been most appreciated.

Emma Mosley should not be forgotten for she was the key to solve administrative puzzles.

I want to emphasise the warm welcome I received from the team in Egham.

I have a special thank for Professor David Naccache, who helped make this PhD research possible.

Contents

List of Figures

List of Tables

List of Notation

A	Authority
T	Terminal
SC	Smart Card
SK_Z	Private key of the entity Z
PK_Z	Public key of the entity Z
K	192 bits AES session key
K_{MAC}	AES key of 192 bits which will be used to generate a MAC
SV	Starting Variable for the CBC mode of encryption.
SV_{MAC}	Starting Variable for CBC-MAC generation.
$S_{SK_Z}\{X\}$	Element X has been signed with the private key of the entity Z
$e_Y(X)$	Message X has been encrypted by an asymmetric encryption method using the public key Y
$E_Y\{X\}$	Message X has been encrypted by a symmetric encryption method (AES using the CBC mode) using key Y
$MAC_Y < X >$	MAC computed on the data X using a CBC-MAC method with key Y
ID_Z	Identification number of entity Z
n_{Z_i}	A nonce generated by entity Z. i Stand for the i^{th} generated nonce.
$Gesture_i$	Fraction of the gesture, where i stands for the i^{th} fraction of gesture

List of Abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
AES	Advanced Encryption Standard
APDU	Application Protocol Data Unit
ATM	Automated Teller Machine
CBC	Cipher Block Chaining
cm	Centimetre
DNA	DeoxyriboNucleic Acid
DoS	Denial Of Service
DTW	Dynamic Time Warping
EBGM	Elastic Bunch Graph Matching
ECG	ElectroCardioGram
EEG	ElectroEncephaloGram
EER	Equal Error Rate
FA-AK	False Acceptance attacker knows
FA-BF	False Acceptance brute force
FAR	False Acceptance Rate
FLDA	Fisher Linear Discriminant Analysis
FRR	False Reject Rate
GB	GigaByte
GDA	General Discriminant Analysis
GR	Gesture Recognition
GSM	Global System for Mobile
HCE	Host Card Emulation
IBM	International Business Machines
ICA	Independent Component Analysis
ID	Identification
IR	InfraRed
IoT	Internet of Things
IT	Information Technology
LDA	Linear Discriminant Analysis
LED	Light Emitting Diode
LFA	Local Feature Analysis
LH	Left Hand

LIDAR	Light Detection And Ranging
MAC	Message Authentication Code
m	Meter
NIR	Near InfraRed
NSTC	National Science and Technology Council
OTP	One Time Password
PCA	Principal Component Analysis
PDA	Personal Digital Assistant
PIN	Personal Identification Number
PKI	Public Key Infrastructure
RAM	Random Access Memory
RFID	Radio Frequency IDentification
RH	Right Hand
ROC	Receiver Operating Characteristic
RSA	Rivest, Shamir & Adleman (public key encryption technology)
SC	Smart Card
SE	Secure Element
SIM	Subscriber Identity Module
SK	Private Key
SMS	Short Message System
TAR	True Accept Rate
TEE	Trusted Execution Environment
TRR	True Reject Rate
USB	Universal Serial Bus

List of Publications

1 Benoit Ducray, Sheila Cobourne, Keith Mayes and Konstantinos Markantonakis, "Authentication based on a changeable biometric using gesture recognition with the kinect$^{\text{TM}}$." In 2015 *International Conference on Biometrics* (ICB), pages 3845. IEEE, 2015.

2 Benoit Ducray, Sheila Cobourne, Keith Mayes and Konstantinos Markantonakis, "Gesture Recognition Implemented on a Personal Limited Device." In *The International Conference on Information and Communication Systems* (ICICS), 2017 8th International Conference on, pages 171-176. IEEE, 2017
(Nominated for the Conference's Best Paper Award)

3 Benoit Ducray, Sheila Cobourne, Keith Mayes and Konstantinos Markantonakis, "Comparison of Dynamic Biometric Security Characteristics against other Biometrics", To appear in the IEEE ICC 2017, *IEEE International Conference on Communications*, Paris, France, May 21-25 2017

Chapter 1

Introduction

Contents

This chapter presents the motivation for the thesis. It shows that all authentication factors have limits and that single-factor authentication is not secure.

We introduce the main contribution, which proposes an authentication system that is secure, easily changeable, and multi-factor.

The chapter concludes with the presentation of the structure of the thesis.

1.1 Introduction

The explosive growth of the consumer electronics, communication and the use of internet has given rise to an impressive amount of information which needs to be stored, processed and communicated securely.

Attempted violations of privacy, confidentiality, authority, access and control of the information, for either mischievous or vicious purposes, are commonplace. This generates a powerful need to efficiently protect access to certain resources, reserving them for those who are authorised.

Recognizing authorised individuals requires auditing their claimed identity. Methods to authenticate identity are based on one or more of the authentication factors that were formalized in the 1960s by IBM [112], namely:

- The presentation of *something that one has*. This currently involves various forms of security token such as password generators, Smart Cards (SC) or USB keys, which, when presented to the system allow access. This is known as the "ownership factor".
- *Something that one knows*, which these days commonly takes the form of a PIN or password, among other things. This is known as the "knowledge factor".
- Demonstrate *Something that one is/does* which nowadays breaks down into various electronic techniques of biometric measurements (physical, physiological and behavioural) and gesture recognition. This is known as the "inherence factor".

Each of the above factors has limitations and weaknesses in terms of security and usability, which can be exploited by potential intruders.

Both "something you know" and "something you have" suffer from weaknesses and inconvenience. Things you know can be forgotten or discovered by a fraudster; things you have can be lost, stolen or counterfeited. These nuisances have led to increased use of biometric authentication as a means of increasing security. However, conventional biometrics are static, so if compromised cannot be changed by the user. This has led to interest in techniques that authenticate using changeable multi-factor authentication measures that are influenced by biometrics, rather than completely reliant on them. Investigating the practicality and security of such techniques provided the motivation for this research.

1.2 Contribution

First we needed to identify “easily changeable biometrics” which led us to define a new biometric family: **the dynamic biometric**. This new family has some security characteristics which we studied and compare to several fixed biometrics based on criteria devised by Bonneau et al. [23].

Out of the dynamic biometric, we chose to focus on one particular type which has become possible through the popularization of depth sensors: **the authentication based on gesture recognition**.

We focused on assessing if a Gesture Recognition (GR) application would be feasible and secure.

To do that we conducted several experiments in order to determine if GR would be able to authenticate genuine users and reject any impostors. This was tested when these impostors do not know the gesture and do random movements, and when they do know the gesture and try to mimic it.

We then put our results into perspective by comparing them with others works. We broadened the analysis to another widely used biometric, fingerprints, thus giving a point of comparison to other form of biometrics.

We continued by looking for a secure place to store the Gesture Recognition’s template and run the authentication application. We evaluated the possibility of doing that on a personal limited device, such as Smart Card (SC). The result of this experiment shows us that on a SC the application would require more than a minute to run but if implemented using Host Card Emulation (HCE) on a mobile phone it would not need more than 1 or 2 seconds.

We designed a protocol to use a Gesture Recognition application which we assess against several threat vectors defined by Roberts [94].

1.3 Structure of the thesis

To efficiently conduct this research work, we relied on two complementary approaches, namely performing a review of the literature, and conducting several experiments.

The thesis is presented in three parts: part 1 “Background”, part 2 “Contributions” and part 3 “Conclusion”. The “Background” consists of chapters 2, 3 and 4 which include of literature searches relating to the areas of individual authentication (Chapter 2); existing biometric techniques, modalities and threat vectors used for such purpose (Chapter 3); various methods used for gesture recognition and tracking in order to make some kind of inventory, understand the concepts, approaches and techniques used (Chapter 4).

In the part named "*Contributions*", we gathered all our research and experiments as well as their analysis.

We first set the definition of what we named a "dynamic biometric" in Chapter 5 entitled "**Comparison of Dynamic Biometric Security Characteristics against other Biometrics**". In this chapter we also studied the security properties of dynamic biometrics and compared them to those of other biometrics.

Then, we examined the feasibility and the security of GR to authenticate genuine users and reject any impostors when these impostors either know or do not know the gesture.

In Chapter 6 titled "**Feasibility of authentication based gesture recognition**". We conducted several experiments with different sensors where we imposed the gestures or let the volunteers to choose them. We compared our results to others' work done on GR and went a step further by comparing them to a widely used biometric, namely fingerprint recognition.

In Chapter 7 called "**Gesture Recognition Implemented on a Personal Limited Device**", we investigated storing and protecting a template and running a Gesture Recognition application on a personal limited device such as a SC or on a smartphone by using an HCE application.

Chapter 8 named "**Application and threat model**" presents a protocol for using an authentication based on GR application. This chapter also presents an analysis of the security of this application against different threat vectors and how various profiles of attacker may take advantage.

In the third part we will present the conclusion of our research, as well as some proposal for future research work in the field.

Part I

Background

Chapter 2

Authentication

Contents

This chapter presents the authentication factors, i.e. the knowledge factor, which relates to "something the entity knows"; the ownership factor, relating to "something the entity has/possesses"; the inherent factor, which can be summarised as "something the entity is or does", as well as their use in our everyday lives.

This chapter also gives the definition of single- and multi-factor authentication and it enlightens the context in which each of them should be used.

2.1 Introduction to authentication

The multiplicity of services and functions we can access on-line today has changed drastically in the last decade, both in the way we interact with others and use those technologies.

In the past, people had to rely on physical relationships and personal trust, whether it was for their personal matter or business, whereas nowadays they are largely and increasingly facing machines on which processes for many activities have been automated.

Even with this physical relationship, the recognition of the right person was sometimes tricky as [13] reminds us using a Bible story (Genesis, 27:1-24). With virtual communications expanding and the erosion of personal contacts, authentication of individuals has become even more difficult in the on-line world.

Some consumers may not feel confident with on-line technologies and are reluctant to make use of them to purchase goods even though prices can be very attractive in comparison with retail offers. On the contrary, others are more relaxed and have accepted these new ways, providing their banking details as they would have done over the counter of the corner shop. Sometimes this trust can be misplaced.

The expanding use of web-based technologies has been helped by a number of factors.
For example, the fact that government agencies are digitizing information and proposing more and more of their services on-line, means citizens may prefer to remain comfortably at home and access the services on-the-spot instead of going "there" and queuing (sometimes for hours) before being served.
However, this ease of use also helps the development of identity theft with its heavy financial and social burden [103].

It can be seen that there is a necessity for systems to allow reliable and strong validation and verification of the identity for both parts of an exchange [77], at both national and organizational levels.

This has been called "Authentication".
This is the process that aims at confirming the claimed identity of an entity (person or computer). Depending on the result of this authentication process, access to the required resource (network, application...) may be granted.

In our context, at the beginning of the 21^{st} century, with an increasing presence of social networks, the development of IoT and the robotisation of more and more services, authentication has become a crucial function and it is the key concept within which our work on the different projects has been accomplished.

2.2 Authentication factors

2.2.1 The three main authentication factors

Over the years and centuries, people have imagined and created several means to segregate who was part of the "allowed ones" from everyone else.

The most natural way to identify a person, and the one we use in our everyday relationship with the people "we know" such as our friends, family, neighbours and colleagues, is to rely on their aspect and appearance, i.e. the characteristics that qualify the expression "something one is".

Historically, there are numerous examples of people having been deceived in such manner. For example, in the book of Genesis there is an example of this when Jacob pretends to be Esau, Isaac's first-born son. Despite the doubts Isaac has because the voice speaking to him seems to be Jacob's, Isaac tries to identify Esau by bringing him back some game for his meal, touching his hands and arms in search of the hairy feeling as a characteristic.

As relying on the aspect and appearance of somebody is not infallible and because one cannot know a large amount of people coming from elsewhere, another approach has been conceived.

The term "password" refers originally to the act of being requested to pronounce the expected word(s) to be recognized and then granted access to a given place.
A different password might be required if the individual wanted to enter a more profound circle. Security was then based on "something one knows".

On another hand, some very early people imagined a different way to authenticate, using what the ancient Greeks called "syn-bolun" (which has become "symbol"). For example, part of a given object which could be compared or assembled with another one that the counterpart would have had in his possession in order to assess the validity of the identity of both persons. Security was then based on "something one has or possesses".

It might even happen than in order to get more security, several of these previous techniques were used at the same time, such as the combination of a physical symbol and a password.

With the augmentation of digital information as well as the number of people who can potentially access it, early in the story of the computer, from the end of the 60's [112] IBM defined the three authentication factors:

- **Knowledge factor**, which relates to "something the entity knows", for example a password, a personal identification number (PIN).
- **Ownership factor**, relating to "something the entity has/possesses", for example ID card, software and/or physical token.
- **Inherence factor** which can be summarized as "something the entity is or does", for example a signature, a gesture or all kinds of biometrics.

Some authors are researching other authentication factors than those listed above: for example "where the entity is located" [13] which can make sense for specific applications that can only be accessed from pre-determined terminals (more and more difficuly in a nomade era), or "someone you know", i.e. having a reference person that could help and grant urgent access in case of problems [24].

But these two examples can only be exceptional processes as treating a large crowd in such ways would be rather difficult and expensive.

2.2.2 Knowledge factor

Knowledge authentication factors have at their base a shared secret. The principle is that both parties, viz. the user (something a person knows) and the entity to whom the user has to authenticate, know a specific piece of information.

Passwords and PINs are the most common examples of this type of factor.
This can also include questions or queries which should only be answerable with the user's specific knowledge, as well as selecting, recognizing or identifying pre-chosen (during initial enrolment process) images mixed in a pool of images [34].

The security level is largely affected by the lifetime of this secret.
If the shared secret never (or scarcely) changes over time, it then would be described as **"static"** (or "fixed") and the risk for this secret to be compromised increases as the time passes-by.
A secret with a short lifetime will be more difficult to discover before it is obsolete.

2.2.3 Ownership factor

Ownership factors (something a person has) use tokens. These tokens are physical devices providing all or part of the information required by the authentication authority. There are numerous kinds of token, such as an ID card, a passport or a driving license.

The three following tokens are currently the most secure and widespread: the USB token device, the smart card, and the password-generating token [34].

The USB token device

The USB token is a device which is quite easy to carry and user-friendly.
With its relatively small size, as well as the way it works, it can be compared to a house key. The user just has to plug it into a computer's USB port, like a key within a lock.
Once the machine recognizes the token, the owner gains access either to a password prompt or to the computer system.

The internal memory of the USB device stores, solidly encrypted, the required programs certificates and encryption.
No installation of any extra hardware on the computer is required as the system uses the existing USB port as long as there is one (not all new computers have standard size USB ports, or any at all).

If the USB drive is based around an attack-resistant chip and/or enclosure then it can be used in a public key infrastructure (PKI) environment.

The smart card

A smart card (SC) consists of a tamper-resistant integrated circuit embedded within a card carrier [44].

It has to respect international standards and its maximum size should not exceed that of a credit card and is often smaller as in the case of the Subscriber Identity Module (SIM) cards used in mobile phones.

Strictly speaking a smart card should have a card package, however the terminology is often used for other packages/carriers such as tags, key fobs, watches, USB tokens etc. [68].

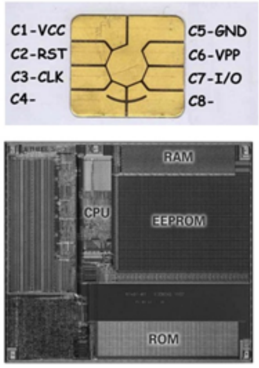


Figure 2.1: *Contact Smart Card under a microscope (source [68]).*

The use of the denomination “smart” is explained by [68] as follows:
“The unique ability to store relatively large amounts of data, carry out their own on-card functions (e.g. encryption), interact intelligently with a smart card reader, with the help of an embedded microcontroller and support mutual authentication” (using for example the two unmarked components shown on Figure 2.1).

Most importantly, the SC is able to resist attacks against its normal operation.

The work done by [68] also explains that a typical smart card system is composed of the cards, the readers and the background system, the whole being described as follows:
The use of an SC requires a reader to which it connects either with direct physical contact (called a ‘contact card’) or with a remote contactless radio frequency interface (called a ‘contactless card’ which can be broken down into ‘proximity cards’: distance not exceeding 10cm, and ‘vicinity cards’ distance not exceeding 1.2m [44]). The reader connects to the background system which stores and processes the information of the whole system.
When the SC is recognized as valid (ownership factor), the user is allowed to enter a password (knowledge factor).

It is explained in [44] that although contactless smart cards use radio frequency fields for their communications (and usually – not always – as a source of power), they are different from RFID devices that are not restricted to card carriers and can be embedded into a range of objects. RFID is not as sophisticated as contactless smart cards, not because of the technical limitations but due to functional and cost requirements. “RFID refers to procedures to automatically identify objects using radio waves” [44].

Mayes et al. explain in [77] that since SC can store secret identifiers securely and engage in cryptographically protected (challenge-response) protocols, it is generally accepted that SC currently play a very useful role in secure authentication.
They may contain some personal information about the user, like banking cards, strongly linking them to a personal identity, or not as it is the case of telecommunication SC, which often have only a weak linkage to the person with the emphasis being more on a valid (and paying) account.
In either case, the authors emphasize that the use of a SC exploits high security protocols and processes in a very user friendly way that provides both “easy-to-use and tamper-resistant security”.

All the SC advantages that are recognized by [34] are the same as those described above for USB tokens, but from these authors' point of view, the disadvantage of the SC lies in the need for the users to install additional hardware, i.e. the reader, as well as associated software drivers on their own computer.

Password-generating token

The principle of a password-generating token is to generate a One-Time Password (OTP) which would be displayed on the screen embedded in the token. This passcode is unique and either changes after a certain amount of time (some may last 60 seconds and others may last 30 seconds) or once it has been used.
The system ensures that the same OTP cannot be used twice consecutively.

In the case of the use of a token continuously showing an OTP which changes regularly, when accessing an information, the user might be required to enter their username as well as a standard password (knowledge factors) which are completed by the OTP generated by token (the ownership factors). The condition for the user to be authenticated is the match between the authentication authority's expectations and both their password and the OTP.
This kind of authentication is considered secure by [34] because in a time sensitive context the OTP are unique, random and unpredictable.

Non-hardware-based OTP scratch cards are less-expensive, "low-tech" versions of the OTP generating tokens discussed above. It is a card, similar to a bingo card in the sense that there are cells (the number of which depends on the physical size of the card) presented in a grid which contains alphanumerical characters.

According to [34], if these scratch cards are placed inside protecting plastic that makes it durable and easy to carry. They present some advantages over conventional OTP hardware tokens which rely on electronics that can fail through physical abuse or defects.

This type of authentication has the advantage of not needing any training and in the case where the card would be lost, the replacement of this one is relatively easy and inexpensive.

Another authentication technique is to send an SMS (Short Message System) with a code on the user's (known) mobile phone and asking them to enter that code.
Such technique transforms the mobile phone into a temporary token.

2.2.4 The inherence factor

The etymology of the term ‘Biometrics’ is literally measurement (metrics) of the living (bio). It refers to identifying a person based on one or more of their physiological or behavioural characteristics [82].
In other words, the identification is proceeded based on “something a person is or does”.

Any physiological and/or behavioural characteristic of a person that satisfies the distinctiveness, permanence, universality and collectability properties can be used as a biometric.

Physiological, physical and/or anatomical characteristics relate to measurable biological specific parts of a person (e.g., face, speech, fingerprint, iris). They are related to the shape of a part of the body.

Behavioural characteristics (e.g., gait or speech too) are related to the behaviour of the person and are the result of an acquisition/learning process. They are thus comparably less stable as they might evolve over time because of external influences.

The work of [97] describes how physiological characteristics are generally more stable over time than the behavioural.

2.3 Single- vs multi-factor authentication

Several types of factor-based authentication are important to differentiate when one wants to deal with secure authentication.

2.3.1 Single-factor authentication

A system which uses only one of the above three factors of authentication is called “single-factor authentication”. The most common example of this type is the use of a password to authenticate a person.

Single-factor authentication as the only control mechanism is considered by the US federal financial agencies (e.g., the Board of Governors of the Federal Reserve System, the Federal Deposit Insurance Corporation, the National Credit Union Administration, the Office of the Comptroller of the Currency, and the Office of Thrift Supervision) to be inadequate for high-risk transactions involving the movement of funds to other parties or access to customer information [34].

2.3.2 Multi-factor authentication

A system which uses two of the above three factors of authentication is called “two-factor authentication”. A example of this case is when a user gets money from an Automated Teller Machine (ATM), they both need to have the bankcard (ownership factor) and to know the PIN (knowledge factor).

In the same way, three-factor authentication is the use of all three of the factors of authentication. For example, in order access a highly secure site an individual might need to pass a guard who checks their face against a stored image (something you are), use an access card (something you have), and enter a code (something you know).

In a more general way, the use of multi-factor authentication mechanisms refers to the application at least two authentication factors. This allows for reliable and strong authentication which can be more difficult to compromise due to the fact that multiple shared secrets (more than one) must be known to authenticate [34].
It is preferable to use multi-factor authentication to protect high-risk, high-stake information.

Computer security systems and technology have passed through several changes. According to [115], “the trends have been from what you know (e.g., password, PIN) to what you have (ATM card, driving license, etc.) and presently to who you are (biometry) or combinations of two or more of the trios”.

A multi-factor authentication system is, from a security point of view, more difficult to defeat as it would require the attacker to discover/replicate/obtain several things such as the shared secret (something you know) as well as the token (something you have) and/or replicate what you are (biometrics).
The literature states that any authentication factor can potentially be defeated but to deceive a multi-factor authentication system it would involve more effort and money.
Multi-factor authentication is however sometimes perceived by users as inconvenient and slow.

Chapter 3

Biometry

Contents

This chapter will detail further what relates to biometrics in order to allow for a good understanding of the subject.

After a quick overview of the definition and some clarifications about the vocabulary used in the present research.
We will first present the processes of biometric-based systems, and then an overview of the main techniques that are today either in use or in research.

3.1 Definition and successful evaluation

3.1.1 General principle

As PINs and passwords (something the user knows) can be either learnt by an attacker and/or forgotten by the user; tokens (something the user has) can be stolen or lost. All of this increased the interest in biometric solutions (something the user is) [77].
These are ways to identify a person based on either their behavioural or physiological characteristic as such techniques have the ability to reliably distinguish between an imposter and an authorized person [82].

An intrinsic advantage is offered by biometric authentication which is based on direct evidence of the personal identity (whereas secrets of any type can be stolen) when compared to other non-biometric identifiers, especially when considering that biometry enrolment is made in a controlled environment making it very reliable for future use [103].

However, to be more precise, as [22] said, biometrics describe both behavioural characteristic(s) of a person and/or measurable biological (physiological and anatomical) as well as the process (developed later) of using automated methods of verifying, identifying or recognizing an individual based on such characteristic(s).

Recognition, verification, identification, authentication

“Recognition” is a generic term which does not necessarily imply either identification or verification.

Verifying or identifying a person’s identity are two concepts which are not exactly the same from a biometric process point of view [57, 77, 22, 112, 82]:

- To verify the identity, the biometric data of the person is captured and compared with the biometric data of this person which has been previously recorded during enrolment (see later) and stored in a database.
  This is a one-to-one (1:1) matching.
  In this case a person submits a claim which will be verified and after verification is either accepted or rejected.

- To identify a person, the biometric data of the person is captured and compared with a whole set of biometric data available in a database in order to identify (i.e. give an identity to) the person from whom biometric data was captured.
  This is a one-to-many process (1:n, with n being the size of the database).

So in this case a specific identity of an individual is identified without submitting a claim [82].

3.1.2 Components and processes involved

A 3-step process

A recognition system based on biometrics will operate as follows:

- Acquisition of the biometric data from an individual (capture);
- Preparation of the biometric data template by extracting feature sets;
- Comparison between the extracted biometric data and the reference template which is in the database.

"Enrolment" [60] is the process during which the biometric data of an individual user is captured. The features are extracted in order to save the template in the database of the authentication system as a data sample from one or more physiological or behavioural characteristics.

Generally biometric enrolment is run in-person and in a controlled environment that aims at making it very reliable for future use [103].

Presentation of the components of a biometric system

A typical biometric system is comprised of five integrated components [22, 43]:

1. A sensor is used to collect the data and convert the information to a digital format. Sensors involve amongst other things: cameras, microphones, fingerprint scanners, etc.

   Each digitization of the original biometric signal potentially creates what is called 'noise' (i.e. unwanted components in a signal) which distorts the signal-image.

   The quality of the capture steps is crucial for the whole system efficiency.

   It relies not only on technological elements (sensor, digital template) but also on the controlled conditions and environment of the enrolment (lighting, background noise, etc.).

   Two samples captured at different times will not be exactly equivalent as some variation in presentation will occur.

2. The digital signal-image is then transformed into a digital set (called a template) through signal processing algorithms which perform quality control activities and develop the

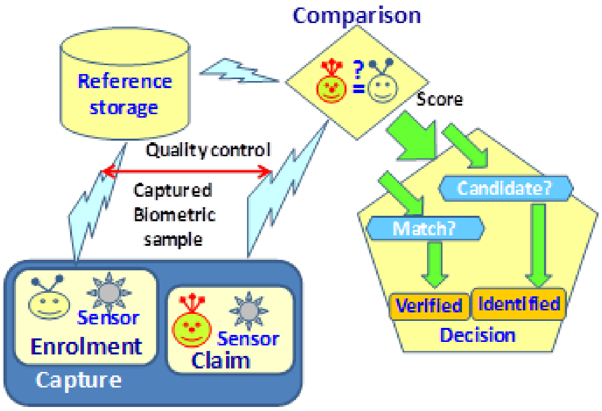


Figure 3.1: *Biometric authentication system (image adapted from [43])*

biometric template, eliminating the noise as well as unnecessary information (such as background).

Templates will be used as references in the following steps of the process.

Their quality is therefore important and it may be necessary to make several attempts to capture before a reference template is created.

3. Template data will be encrypted for the sake of security and privacy and stored either in a database (if remote, the transport network needs to be secured because the template could be either exchanged or altered) or on a card for later comparison with an identity claim data set.

4. When this occurs, a specific algorithm:

   - compares the claim template to one or more templates kept in data storage
   - and gives the matching score that indicates the degree of similarity, i.e. the likelihood that both data sets are coming from the same individual.

5. Finally, a decision process (automated or human-based) makes a decision based on the comparison of the matching component results (i.e. the score) to an acceptance threshold (called ‘match threshold’) that has been previously defined according to the system’s sensitivity.

Figure 3.1 illustrates the various components of a biometric authentication system, as well as the different steps of the process.

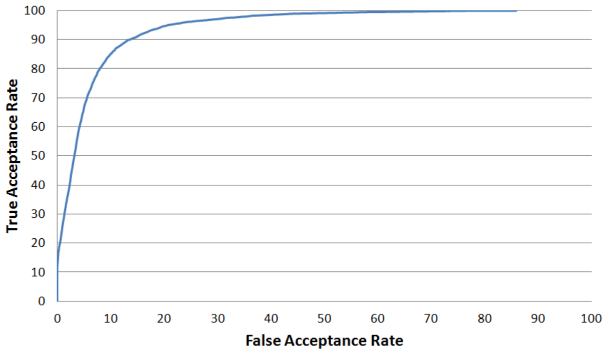


Figure 3.2: *Receiver operating characteristic (ROC) graph (source: Benoit Ducray)*

3.1.3 Match between threshold and management indicators

Human recognition systems are inherently probabilistic, and hence inherently fallible.
The chance of error can be made small but cannot be eliminated. System designers and operators should anticipate and plan for the occurrence of errors, even if errors are expected to be infrequent [32].

Biometric matching is probabilistic in nature, which implies that two samples of the same individual are never exactly the same [103] for multiple reasons including environmental conditions, sensibility of the sensors, reaction/adaptation of the user to the sensors and process. Hence those systems are inherently fallible.

The developers who build the application know (or should know) the quality, capability, and sensitivity of the components they use, so they can specify what is called a “match threshold” that defines the level of concordance required for acceptance.

If technically possible, setting the required score at 100% would be ideal.
But it would also cause inconvenience to users who would be required to be very thorough when going through the authentication process in order to have a chance of being validated.

The match threshold can be defined based on the Receiver Operating Characteristic (ROC) approach which is a method of showing measured accuracy performance of a biometric system.

A verification ROC (as shown in Figure 3.2) compares false accept rate versus verification rate whereas an open-set identification (otherwise called ‘watchlist’ ROC) compares false alarm rates against detection and identification rate.

The statement presented in Table 3.1 shows the various output triggered by the comparison

Table 3.1: *Matching rate for acceptance (source: Benoit Ducray)*

Score level	**Genuine sample from Legitimate owner**	**Fake sample from impostor**
Superior or equal to threshold	Correctly accepted	Incorrectly accepted
Management indicator	TAR (True Accept Rate) Verification/Identification Rate	Error type II : FAR (False Acceptance Rate)
Inferior to threshold	Incorrectly rejected	Correctly rejected
Management indicator	Error type I : FRR (False Rejection Rate)	TRR (True Reject Rate)
None	Failure to acquire. New capture required	
Error types	FTA (Failure to Acquire), i.e. failure to capture/extract usable information from a biometric sample.	
	FTE (Failure to Enrol), i.e. failure to form a proper enrolment reference for a given user due either to a lack of training in order to provide their biometrics, or to the quality of the sensor(s) that are not capturing data correctly or with the expected quality level to develop a template.	

of the score with the match threshold as well as the management indicators and error types that are related to each situation.

3.1.4 Multimodal biometrics

Multimodal biometric systems refer to systems that use more than one biometric characteristic to authenticate individuals [43], whether several instances of the same modality (e.g., several fingers or both eyes) or different modalities (such as fingerprint and eye recognition).

Figure 3.3 shows the various possible combinations of a multimodal system.

In the same way that multi-factor authentication offers increased security, the biometric multimodal approach drastically improves recognition accuracy and reliability [96].
Each biometric modality of the solution brings additional (fairly) independent pieces of evidence and authentication is based on global consolidation.

This can be useful for both increased authentication (a single biometric is not so simple to

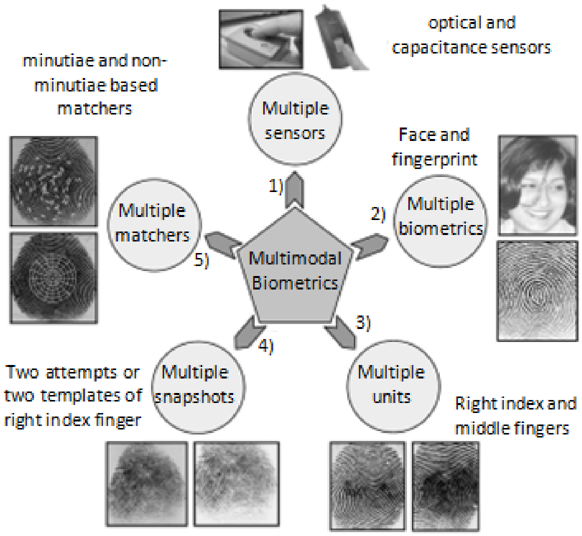

optical and capacitance sensors
minutiae and non-minutiae based matchers
Multiple sensors
Face and fingerprint
1)
2)
Multiple biometrics
Multiple matchers
5)
Multimodal Biometrics
3)
4)
Two attempts or two templates of right index finger
Multiple snapshots
Multiple units
Right index and middle fingers

Figure 3.3: *Multimodal biometrics scenarios (adapted from [96])*

forge but several can become even more difficult for fraudulent identification) and in situations where a given biometric is not shared by all users.

Such an approach can also be used to mitigate risks and diminish error rate, by bringing some solution to both poor sensor performance (as each capture will be scored and the authentication will be granted on the basis of the fusion of all scores), as well as for ensuring aliveness of the captured sample [96, 43].

Given the various modalities used at the same time in multimodal biometric system, information analysis requires reconciling/combining the different elements collected in order to have one single result at the end of the process.

Reconciliation can potentially occur at any of the three steps of the process [96], i.e. fusion can be made at:

- Data or feature level, which means that either the data itself or the feature sets originating from multiple sensors/sources are fused.

  Fusion at that level theoretically brings better recognition results but in practice it is difficult to achieve, especially because the feature sets of the various modalities are often not compatible and/or are not made accessible by equipment providers.

- Match score level, in which the different scores obtained by the multiple classifiers of the different modalities are combined.

This is the commonly preferred solution, as the scores resulting from the different modalities are available and relatively easy to combine.

- Decision level, which means that the final outputs of the different classifiers are consolidated (sometimes using something close to majority voting). This is often considered to be difficult to use because of the limited amount of information available.

However, several studies have questioned the robustness of multimodal systems [95, 9, 10]. Rodrigues et al. in [95] evaluated a multimodal system composed of face and fingerprint using three different fusion schemes: weighted sum, likelihood ratio and Bayesian likelihood ratio. They found that the multimodal systems have a very high probability of being spoofed when only one of its modes is spoofed. Different fusion schemes were tested by Akhtar et al. in [10]: product, perceptron-based rule, etc. They experiment shown that multimodal biometric systems are not intrinsically robust against spoof attacks.

3.2 Overview of biometric techniques / modalities

The following is a quick overview of different biometric modalities in order to help understand what they are about as well as their diversity and potential limits. They are presented sorted into three families which are physical, behavioural and physiological.

3.2.1 Physical biometrics

Face

Face recognition technique is non-intrusive and easy to use, especially as numerous cameras can catch the required data, which makes it one of the most popular biometrics.

It is the preferred way of humans to recognize each other but a facial recognition system needs to overcome a three-step challenge as it should automatically:

- detect whether there is a face in the image captured by the system,
- locate the face if there is one,
- recognize the face from a general viewpoint (i.e. from any pose, which means whether smiling or frowning).

The best known methods for automated face recognition are based on either facial attributes' (i.e. the eyes, eyebrows, nose, lips, or chin) location and shape as well as their spatial relationships, or the global analysis of the picture of a face as a weighted combination of several reference faces.

Face recognition authentication systems sometimes require respecting some constraints related to the composition of the background and/or illumination conditions.

Jain et al. highlight in [59] that it is questionable whether the face itself, with no contextual information, is a sufficient basis for identifying a person from a large number of identities with a high level of confidence.

Face recognition from 2D images proves to be difficult because of the pose, expression and illumination of the face. Each of these things generate important statistical differences from one to the other.

Using a 3D image makes it easier to localise face features without pose and/or illumination problems [82].

Numerous algorithms have been proposed to improve the face recognition over the years.

Both [82] as well as [22] propose to class them into two families which are:

1. Appearance-based methods, such as the following:

   - Fisherfaces [20] which are also called Linear Discriminant Analysis (LDA) or Fisher Linear Discriminant Analysis (FLDA)
   - General Discriminant Analysis (GDA) [19]
   - Eigenfaces [110], also known as Principal Components Analysis (PCA)
   - Neural Networks [67]
   - Independent Component Analysis (ICA) [18]

2. Geometry feature-based methods, such as the following:

   - Active Shape Model [33, 118]
   - Local Feature Analysis (LFA) [88]
   - Elastic Bunch Graph Matching (EBGM) [116]

A disadvantage of appearance-based methods is that successful face recognition requires having a similar lighting and pose reference in the database.

As a matter of fact it tends to be difficult for these methods to match face images showing either two really different viewpoints or lighting conditions (for example, outdoor lighting versus indoor fluorescents).

Geometric feature-based methods are less sensitive to problems of viewpoints and lighting variations but are weaker in the feature extraction process [82].

This might explain why, according to [22], the three predominant approaches are the following:

- Fisherfaces require knowledge of both the within-class (i.e. within user) as well as the global variations. Thus, this approach requires having multiple samples of each person in the database.
  It is a statistical approach for classifying samples of unknown classes based on training samples with known classes which maximizes the ratio of 'between-class' (i.e. across users) scatter to that of 'within-class' (i.e. within user) scatter.

- Eigenfaces or PCA, the process of which is described by [22] as follows:

  First, the probe and gallery images must be the same size. The probe has to be standardized to line up the eyes and mouth of the subjects within the images.

  The eigenfaces, which are the orthogonal decomposition of the facial patterns are obtained using the PCA approach, are then used to remove information that is not useful. These are then used to keep the most effective low dimensional structure of the face (this also reduces the dimension of the data).

  Each face image can be summarized as a weighted sum (feature vector) of the eigenfaces. When required, the probe image is compared to the gallery image by measuring the distance between their respective feature vectors.

  With the PCA approach, the full frontal face must be presented each time an authentication is required. Other viewpoints result in poor performance [22].

  Another drawback of this method is that the total scatter across all classes, i.e. all images of all faces are spoiled by a lot of unwanted variations such as variations in lighting or in facial expression.
  Within-class variation (i.e. within user) from lighting and pose (standing straight vs. leaning over) are, most of the time, more important than the normal inter-class variation due to a different identity.

- The EBGM method uses the many non-linear characteristics of the face image that are not addressed by linear analysis methods, such as illumination variations, pose and expression.

  An elastic grid is projected on the face corresponding to a dynamic link architecture issued from a “Gabor wavelet transform”.

  The Gabor jet is a node on the elastic grid [116]. It is the result of a convolution of the image with a Gabor filter, which is used to detect shapes and to extract features using image processing.

  Recognition is based on the similarity of the Gabor filter response at each Gabor node.

  The main difficulty is the need for accurate landmark localization. This is sometimes achieved by combining PCA and LDA methods.

Fingerprint

Fingerprint biometric is the analysis of the ridges and valleys formed on fingertips during foetal development. They are unique for each finger of a person and this is also true for identical twins.

Recognition by fingerprint is a basic method due to its outstanding features of universality, accuracy, uniqueness, permanence and low cost. It is the most popular and reliable technique and is currently the leading biometric technology [58].

According to [74], archaeologists have found evidence that such a technique was used by Assyrians and Chinese for some kind of identification as early as 7000 to 6000 BC.

By introducing in 1880 the use of minutiae features for fingerprint matching, Henry Fauld laid the scientific basis of the modern fingerprint recognition.

Nowadays, fingerprint recognition techniques can be broadly classified as:
- Ridge feature-based,
- Minutiae-based,
- Gradient based [7]
- and Correlation-based [56].

Most automated systems use minutiae points in one way or another to identify a person.

In practice, fingerprint identification is done as follows: minutiae are stored in a template, but only a subset of these has to match for identification or verification.
In most systems, if 10 to 20 minutiae match, the fingerprint is considered a positive match. In today's smart card systems approximately 40 minutiae are stored, because of space restrictions [115].

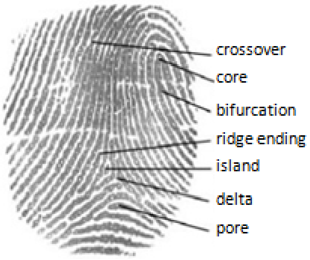


Figure 3.4: *Fingerprint recognition: Examples of specific items creating uniqueness (adapted from [22])*

The matching accuracy using fingerprints, i.e. identification of a person, has been historically shown to be very high, although for a population of more than a few hundred people information of multiple fingerprints of each person are required to allow for large-scale identification involving millions of identities [59].

Fingerprint identification is widely used in personal identification as it works well in most cases.

However, it must be noted that fingerprints of a fraction of a given population such as manual labourers, elderly people, etc., may be unsuitable for automatic identification due to the difficulty to acquire fingerprint features, i.e. minutiae (see illustration 3.4) [59].

This difficulty is a result of various factors impacting the prints themselves or their 'readability'. For example: genetic factors, aging, environmental or occupational reasons (e.g., cuts and bruises on the fingerprints of manual workers make them continuously change).

Palm print

The palm is the region between the wrist and fingers. Palm print features like ridges, singular points, minutia points, principal lines, wrinkles and texture are used for identification.
Ridges and minutiae have a pattern similar to a fingerprint. However, in palm prints the creases and ridges often overlap and cross each other.

Several methods have been proposed for their use in human identification but most can be deceived by forged biometrics.

It is mostly used in combination with hand geometry biometrics [66], which is presented hereafter.

Hand geometry

Another of the most widely used biometric technologies is hand geometry, which becomes stable after a certain age (once an adult).

To be identified, the individual being authenticated just has to show/spread their hand to the sensor. Usually, the process takes only a few seconds for the system to extract features from the captured image and compute the widths and lengths of the fingers at various locations (see Figure 3.5).

As a matter of fact, this modality is based on the analysis and measurement of the overall structure, proportions and shape of the hand, e.g., width, length and thickness of the hand, joints and fingers, as well as characteristics of the skin surface area such as ridges and creases [59].

In order to check for aliveness of the presented hand and to prevent a mould or a cast of it being used, modern systems use several techniques among which are: requiring movement from the fingers, checking skin conductivity, or the heat of the hand.

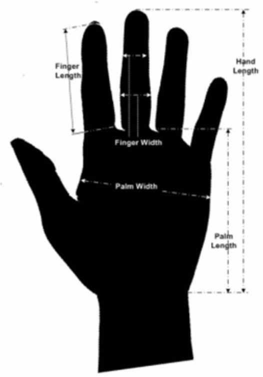


Figure 3.5: *Hand features and geometry measurements used for identification (source [66]).*

As hand biometrics is based on finger and hand geometry, it is also efficient in case of dirty hands, but people with severe arthritis cannot be identified using this modality because they cannot spread their hands on the reader [115].

However, the geometry of the hand is not known to be very distinctive (it seems to suffer limitations that can in a way be compared to Bertillon's anthropometric [48] ones).

It cannot be scaled up for systems used for identification within large populations.

Iris and Retina

Iris recognition is the process of recognizing a person by analysing the random pattern of the iris which is the coloured portion of the eye, i.e. the annular region of the eye bounded by the pupil and the sclera (white of the eye) on either side.
It is a muscle within the eye that regulates the size of the pupil, controlling the amount of light that enters the eye.

Automated methods of iris recognition date back from the mid-1990s.

Although the colouration and structure of the iris is genetically linked, the details of the patterns are not. The iris develops during prenatal growth through a process of tight forming and folding of the tissue membrane.

Using iris patterns as a method to recognize a person dates back to 1936, when ophthalmologist Frank Burch came up with this approach but it was not until 1985 that Drs. Leonard Flom and Aran Safir, also ophthalmologists, proposed the concept that the iris structure is very distinctive and that no two irises are alike.

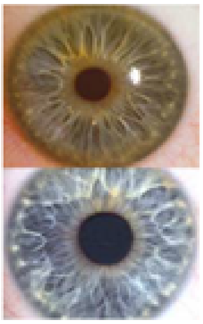


Figure 3.6: *Examples of different iris structures (source [22])*

Dr John Daugman implemented a working automated iris recognition system, which has remained the principal algorithm utilised for that purpose, and which has not had a false match after more than 200 billion comparisons [39].

While iris-based systems show a very low false accept rate (FAR) compared to other biometric recognition, the false reject rate (FRR) of these systems can be high, especially as it requires quite a good user participation [59].

An advantage of this biometric is that it is today extremely difficult to surgically tamper with the texture of the iris and that current systems have the ability to detect artificial irises (e.g., contact lenses).

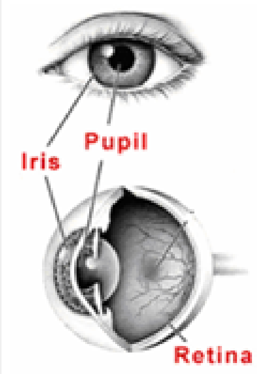


Figure 3.7: *Eye biometric: Differentiating iris and retina (source [22])*

Retinal recognition uses the unique pattern of blood vessels on an individual's retina at the back of the eye. The illustration 3.7 helps clarify the difference between the two organs.

Both techniques involve capturing a high quality picture of the iris or retina using a digital camera. In the acquisition of these images, some form of illumination is necessary.

For that purpose, both techniques use NIR (Near InfraRed) light. Although safe in a properly designed system, eye safety is a major concern for all systems that illuminate the eye.

The risk of damage to the eye is remote with a single LED source using today's LED technology. Illumination from multiple LEDs can, however, produce eye damage if not carefully designed and used [22].

The iris and the retina have higher degrees of distinctiveness than hand or finger geometry [115].

3.2.2 Behavioural biometrics

Signature

This modality should better be called "Dynamic Signature". It does not relate to what the signature looks like, but to the capture through touch sensitive technologies (such as PDAs or tablets) of the direction, stroke, pressure, and shape of a signature. The result can be used as a reliable indicator of an individual's identity.

In practice, signature recognition can only be used for recognition processes rather than identification. It is based on the analysis along the X, Y and Z axis of specific dynamic characteristics such as the speed, acceleration, timing, pressure, angles and direction of the signature's strokes. X and Y positions show velocity modulation in the respective directions while Z indicates changes in pressure with respect to time.

Although quite precise, the characteristics of a signature have a large intra-class variability which means that someone's own signature may vary from one production to the other, often making dynamic signature recognition difficult.

Some dynamic signature recognition algorithms incorporate a learning function to account for the natural changes that may happen over time in an individual's signature [59, 22, 115, 82].

Keystroke dynamics

The use of keystroke dynamics is an automated method of examining the cadence of an individual's typing pattern. The technology uses a specific keyboard compatible with PCs.

It is based on dynamics such as speed and pressure, total time of typing a particular password, and the time a user takes between hitting certain keys.

Keystroke dynamics have the potential for continuous authentication of identity while a person is using a computer [22, 115].

Gait

The 'gait' biometric relates to the analysis of the manner of walking of an individual or more precisely on both static human shape and movements.

It is one of the natural ways commonly used by humans to recognize each other. It could potentially allow for recognition at such distance or at such low resolution that other biometrics might not be perceivable, whether because of distance, concealment or disguise.

Among the behavioural biometrics this is not only one that is quite specific to a given individual but also that might require a lot of training to change as it is something rather natural.

It is based on medical studies and biomechanical literature of beginning of the 90s, but it is still in the research and development stage in order to find an automated way of applying it [115].

3.2.3 Physiological biometrics

Speaker / voice

Speaker or voice recognition uses a person's voice, i.e. the acoustic features of speech that are specific to each individual, in order to recognize or identify someone.
It is important to differentiate the biometric approach, the purpose of which is to recognize or identify an individual, from what is called "speech recognition", which aims to recognize spoken words.

Some authors such as [112] mix the two technologies in their presentation.

This biometric modality is based on the acoustic patterns of the sound of the voice, i.e. the "Voiceprint" building the voice template. These patterns, i.e. voice pitch, speaking style, tone (e.g., low, high, nasal, etc.), cadence and frequency of a person's voice, resonances in the nasal passages, etc., are the result of the combination of two things:

- physical patterns which are inherent to an individual and dependent on the movement of the jaw, tongue, larynx, as well as the size and shape of the throat and mouth;
- behavioural patterns, which are acquired.

From the above one can understand that the identification capabilities of this modality can suffer from sensor quality and "within-class" (i.e. each user) variations.

A speaker recognition system analyses the frequency content of the spoken words and sentences in order to compare their characteristics in terms of duration, intensity dynamics, pitch, quality, and so on.

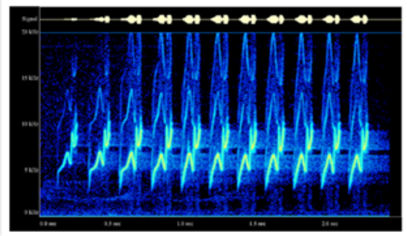


Figure 3.8: *Voice sample spectral analysis of the frequency variation (source [22])*

As shown in the illustration (Figure 3.8), signal speech samples produce waves which can be presented in a kind of graph showing the frequency variation of the voice, with time on the horizontal axis and both intensity and loudness on the vertical access.
Authentication is performed by comparing this graph to the enrolment one and expecting a good match.

Some systems use the “anti-speaker” or “other-speaker” model which compares the captured sample to the template that exists in the database.
The system produces a likelihood ratio (a score) which indicates the probability of a match, i.e. the fact that the presented sample is more similar to the claimed or assumed identity than to the “anti-speaker” model.

Speaker/voice biometric recognition systems can be sorted into two modes of spoken input, i.e. unconstrained (or free) mode and constrained mode.

- The former one (free mode) is text independent which means that any text can be used for the captured sample. In other words, the system does not need the individual submitting the sample to be cooperative or even aware of the capture.
- The latter one (constrained mode) is based on “text dependent” speech which is pre-established. The individual submitting the sample needs to be cooperative as they will be required to read:
  - either a fixed text or passwords. This is called “text dependent”.
  - or a text that is prompted by the system such as “Please say ’36-24-36’ ”.

Prompted text is more difficult to deceive than the two other methods because the text to pronounce is not known in advance (instead, it is automatically chosen by the system) and it changes randomly and completely for every occurrence.

Brainwaves / EEG

A model for using electroencephalogram (EEG) as a biometric identification system has been proposed by Poulos et al. in 1999 and then further explored by various author but this modality is still under research [82].

Heart sound / ECG

According to [82], capturing the heart sound, i.e. the electrocardiogram (ECG) of a person has been found to be quite discriminative and able to allow individual recognition by surveying the anxiety state through ECG features made by [55] in 2005.

It has been shown that these signals are fairly unique to a given individual and not subject to the person's state of anxiety. However, it is a modality still in the research and development stage. Also the many electrodes required by the ECG capture are rather awkward and inconvenient [21].

Vascular Pattern Recognition / Vein Pattern Authentication

Vascular Pattern Recognition, also commonly referred to as Vein Pattern Authentication, is originally based on the work initiated in the 90s by Dr. K. Shimizu [22].

At the beginning of the 21st century, further research has shown that the vascular pattern of the human body is unique for each individual [36] and does not change with a person's age.

By using the difference in near-infrared light absorbance by blood vessels and other tissues, an image of the pattern of the blood vessels (vessel branching points, vessel thickness, and branching angles) of a hand, palm, or finger is produced.

Such systems are currently in use in Japan for securing access to various services, such as ATMs, hospitals, and universities.

DNA

DNA (DeoxyriboNucleic Acid) is present in every living cell and it does not change all through the life of the person.

DNA systems of authentication have a very high degree of accuracy. There are 1-in-6 billion chance of two people having the same DNA profile. Except for monozygotic twins, the DNA system is the most distinct biometric identifier. An individual's DNA profile does not change in the entire life of a person, therefore its permanence is incontestable [115].

Currently the processes to get a DNA sample are rather intrusive and require the person's cooperation. In order to get a DNA profile or DNA fingerprint, it is necessary to get human tissue, blood or some other bodily sample.

Due to the rather long time required for the analysis, this modality is more often used in the forensic side of biometrics.

3.2.4 Other biometrics

This section briefly introduces some biometric modalities which are either in limited use or under development. These are not described in much detail but their existence is important to finalize the review of all the different modalities.

- **Ear-shape**: the shape of a person's ear is known to be distinctive especially in the forensic field.
- **Knuckle-crease**: it seems the image pattern formation where the knuckles or joints of fingers bend is unique enough to enter the realm of biometric identifier.
- **Chemical attributes**: for example, body odour.

The biometrics presented here above are definitely not exhaustive. The review however cover the one that are mostly used and researched.

3.3 Threat vectors

Biometric authentication is part of the whole authentication system and as such is exposed to multiple threat vectors. It is important to consider these vulnerabilities. The aim of these attacks may either be to get authenticated as somebody else or to prevent the authentication of a valid user. The attacker may also intend to prevent access to one or more specific people or deny access to all users.

A threat vector is a route by which the system can be attacked. In [94], the authors list 18 threat vectors, which are illustrated in Figure 3.9.

1. **Denial of Service** (DoS): a DoS attack can be executed through a wide range of routes: e.g., from power loss or physical damage to the system, but all with the aim to corrupt or incapacitate the system. Other forms of attacks could include introducing radio frequency or electrical signals in order to affect data quality.

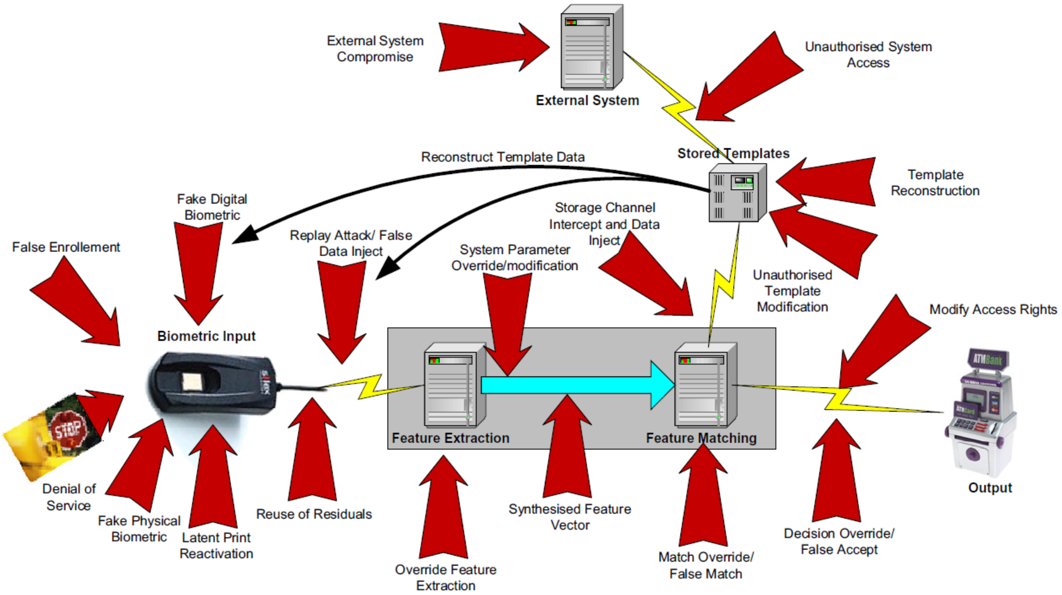


Figure 3.9: *Threat vectors (source [94])*

2. **False enrolment**: the ability of a biometric to authenticate an individual is based on the fact that the enrolled person is who they claim to be. If this identity is a fake, even if the biometric is accurate, the identity will be incorrectly matched.
   In this case the system will validate a false identity, and grant access to privileges.

3. **Fake physical biometric**: this represents a spoofing attack, that is, providing a fake physical biometric in order to circumvent the biometric system.
   These attacks are made on the entry of the system (i.e. the input) so many of the digital protection mechanisms are not effective.

   An original biometric can easily be captured by many sources, with or without the consent of the owner.

4. **Fake digital biometric**: there are two kinds of attack: masquerade attacks and replay of reference sets:

   - Masquerade attacks: this kind of attack uses false biometric data such as digitised latent fingerprints (such as those left on surfaces that have been touched) or digital facial images.
   - A replay of reference sets: this kind of attack takes place inside the biometric system in order to replay the reference set or templates.
     This requires access to the system as well as knowledge of the biometric system.

5. **Latent print reactivation**: some biometrics such as fingerprint and palm print leave a latent print on the biometric sensor due to sweat glands in the skin. A skilled individual may be able to copy or reactivate these into readable prints through a range of techniques.

6. **Reuse of residuals**: some biometric systems and devices may keep in memory the last biometrics extracted and templates used. An attacker who gains access to this data may be able to reuse it to provide a valid biometric.

7. **Replay attacks / false data inject**: this category includes man-in-middle-attacks that consist of capturing the biometrics data and replaying it at a later stage. An alternative way is to inject a false data stream between the processing system and the sensor. That implies, most of the time, some physical tampering with the system.

   A replay attack requires a two or three step process:

   (a) intercept or copy of the sensor transmission

   (b) possibly, modify the data

   (c) replay the data

   An encrypted transmission adds complexity to this process and is an effective defence. Having to decrypt and re-encrypt the data may require advanced technical skills and/or the use of specialised tools.

8. **Synthesised feature vector**: this is an iterative technique which consists of a data stream representing a fake biometric retaining only those changes that improve the score until this fake biometric reaches a point where it gets validated by the system. This technique is described as “hill climbing” and requires access to the matching score.

9. **Override feature extraction**: this attack aims to interfere with the extraction of the biometric features which are being used, either by attacking the software or firmware, in order to manipulate or provide false data.
   This data can then be used to disable the system.

10. **System parameter override / modification**: this attack aims to modify the acceptance threshold or other key system parameters.
    That may result in system acceptance of poor quality or incorrect data.

11. **Match override / false match**: during feature comparison, the templates are generally unencrypted which makes them susceptible to tampering. The decision thus made can be overridden or ignored and modified.

12. **Storage channel intercept and data inject**: if the attacker has access to where the template is stored, they can capture it for later use or inject a false template.

13. **Unauthorised template modification**: the biometric template can be stored in several places such as readers or sensors, on an access card or token, or within the biometric system itself. Any unauthorised changes made as templates are modified, replaced or added to the system are a threat for the system as this may add an unauthorised template or circumvent any registration procedure.
    The loss of template integrity can affect the identification or authentication processes and result into a denial of service.

14. **Template reconstruction**: this attack is similar to synthesised feature vector where "hill climbing" techniques are used, but template reconstitution may also be done by scavenging file fragments from data storage.

15. **Decision override / false accept**: this attack bypasses any identification or authentication processes by overriding the decision data or injecting a false acceptance.

16. **Modify access rights**: this attack is generally achieved by obtaining system administrator rights to allow the unauthorised user to get access privileges and other data and key system parameters.

17. **System interconnections**: the interconnection with other systems may imply two other threat vectors: external system compromise and unauthorised external system access.

3.4 Privacy and biometric authentication

Biometric authentication and the technology progression that enable its application in multiple contexts for both governmental applications (law enforcement, border and immigration management) and private ones (be it health care sector, financial institutions), means that privacy is a legitimate topic [8].

According to [99], security concerns regarding the use of biometrics have increased since 2006 together with and increase in occurrences of fraud. Countermeasures in the form of both algorithms and software have been developed but they are far from being part of standard applications. Some time will be required to secure existing implementation but it is a very important point as any compromised biometrics can have serious consequences for their owner (identity theft being one of them).

Biometric authentication is raising at least three types of issue related to privacy:

- The first one regards the potential access to biometric templates (whether the reference one or the submitted ones) when authentication claims are made, in other words, the privacy and security of the various steps of the authentication process.
- The second type of issue relates to ensuring reference data integrity, which relates to the security and privacy of the technical parts involved in the solution.
- The third one refers to the social, legal and political side of privacy related to the use of the biometric information.

3.4.1 Privacy and biometric submitted claims

The use of fake biometrics, or even stolen ones (e.g., using a photograph for face recognition, forged fingerprints, etc.), by impostors claiming a false identity has not remained in Hollywood movies. It is to present such techniques that aliveness tests are used.

The fact that biometric modalities use real human physiological or behavioural characteristics to authenticate users and that these biometric characteristics are (more or less) permanent and not easily changeable has led people to think (wrongly) that it is quite secure by nature.

In 2006, Jain et al. wrote in [59] that biometrics could not be easily shared, misplaced, or forged and that the resultant security was more reliable than current password systems (which can be shown in hacker websites) with the great advantage not to encumber the end user with remembering ‘long cryptographically strong passwords’.

In addition, biometric traits require the person being authenticated to be present at the time and point of authentication (whereas impostors can negate this if they have access to the secret features).

It is as hard to falsify biometrics, but not impossible, it just takes the right amount of time, money, experience or access.

Nevertheless the biometric data can be stolen from computer systems and networks as well as forged [99] using dead or artificial biometric characteristics. As a matter of fact, biometrics of themselves are not secret, they are a person’s characteristic (we expose them in our everyday life, be it our face, fingerprints, etc.) that can also be captured by malevolent people and then used in place of their genuine owner.

The use of multimodal biometrics as proposed by some authors can help increase the security level but in fact the real problem remains unsolved, and this could create situations comparable to Isaac's one when hearing the voice of Jacob but feeling the hands and arms of Esau and concluding he was in the presence of the latter.

It has been emphasized by [103] that biometric authentication made from a remote location is a clear issue because of the risk of spoofing attacks.

The efficiency of the whole system and particularly the credibility of the score from a biometric matching process depends entirely on the integrity of the sample provided for comparison, and whether it has been provided by the real biometric owner.

In 2009, the NSTC attracted the attention on this point too in showing that, in the context of the US-VISIT application, as well as in other important United States government applications, biometric data is captured in the presence of an officer who can detect the presence of a forgery.

The NSTC added something which is very true and that all users should keep in mind: that once digitized by the system the biometric data is as secure as any other computer data and faces the same vulnerabilities.

3.4.2 Privacy and biometric referential storage

The questions raised by [77] regarding the security of the different steps of the process, the encryption of the data as well as where to store the reference data (centrally or the end-user's smart card) are as important as the technical sides of the rest of the process (capture, quality, transmission, comparison, decision).

If any fraudulent party can access and tamper with the metadata accompanying the biometric data submitted for evaluation, the actual owner of the biometric characteristic can be in serious trouble.

The choice of storing biometric data as well as metadata on a user-held smart card that has the ability to integrate a solution providing tamper-proofed capture, encryption and matching is solving part of the privacy problem (biometric revocation and issue of remote authentication), but not all of it. That SC can still be stolen.

Large centralized biometric template storage raises the question of data integrity all along the process steps as well as the protection of the owner of the biometrics and the revocation of biometrics that have been compromised.

This last point is not an easy one to solve but research has been done on the topic [91, 12, 92].

The risk of forged template or rather of tampering with the template, i.e. changing, transforming or altering the template attached to identification and other metadata, is real and the consequences for the person might be very heavy. The difficulties are compounded because compromised biometric data is very hard to revoke and change (no easy procedure is available for that purpose); this is especially problematic in cases of identity theft.

These problems have been identified by several authors and especially [59].

In this way, [77] emphasizes the fact that choices as to where to store biometric information are limited, e.g., either in a portable device for a small population, or in a centralised database for large scale systems. The latter requires a relatively large, fast and reliable infrastructure especially in order to allow for a continuous service as well as a solution to privacy and portability issues. On the other hand, the authors argue that, using an SC for storing and checking biometric information at the location (match-on-card) allows for more flexibility with a better user-feeling about privacy.

As described above, based on [103] biometric matching is probabilistic in nature, which implies that two samples of the same individual are never exactly the same. The two samples, for obvious security reasons, which are encrypted will need to be decrypted for a comparison to the reference template.
Such a process raises the issue of encryption key management as is the case for any other IT system.

In other words, a biometric authentication system must be considered not just from the technical point of view, i.e. getting the proper biometric data, but more as a complete system including all the management tasks that it triggers, including access to the data and metadata captured and stored as well as modification and update which does not seem to be currently the case when reading the literature we have used as reference for the present document.

As it is said by [59]: “A well-implemented biometrics system with sufficient privacy safeguards may be a clear requirement in the quick response to natural or man-made disasters”.

Such a complete approach is called ‘identity management’.

3.4.3 Social, legal, politics and privacy

In addition to the two previous categories described above, it can be emphasized that the use of biometrics for identity authentication triggers serious questions and even fears in the population.

This is due to the specific nature of biometrics in the sense that they are intrinsically linked to our identity in a way that no other forms of proof of identity (id-card, password, keys, etc.) can approach. Biometrics is so special to so many humans because they are based on physical and/or behavioural characteristics of an individual and because they are very difficult to hide or modify.

The important development of biometric-based identification systems since the 90s which has become quite explosive in the 21^{st} century, especially boosted by the technology. The post 9-11 context has generated a some kind of malaise in the population as Big Brother's technologies were implemented nearly everywhere, leading to what has been perceived as the invasion of individual privacy.

The systematic face recognition scanning of the crowds attending the 2001 football Super Bowl in Tampa, Florida, a few weeks after that September's events, has been one important element to set off the fears in that matter [43].

Civil liberty and consumer groups have reacted to the potential misuse of these data as cross matching with other biometrics databases or sold to third parties. Most countries have then implemented or reinforced a legal framework regulating the use of citizens' personal information (rules of privacy) in a constitutional background, and have defined privacy under law, including physical, decisional, and information privacy [112].

The notion of privacy of the average citizen/user is somewhat proportionate to the quantity and quality of disclosed data required for authentication [103]. One important principle is that biometric information is not to be released without the citizen's prior personal consent.

A side effect of the deployment of biometric technologies has been the difficulties faced by individuals under covert operation or those in witness protection [43].

3.5 Hints for selecting appropriate biometrics

In this section, we introduce criteria that should drive the design of an authentication model based on the use of biometrics.

3.5.1 Apparent advantages of biometric authentication

The chief advantage of biometric authentication methods over other approaches is that they really do what is expected, i.e. authenticate the person actually using the system.

As a matter of fact, all modalities use either physiological or behavioural characteristics to authenticate users which are (more or less) permanent and not easily changeable and therefore the illicit use of somebody's fingerprint or iris pattern does not break security in the same way as any other user's password.

However no technology can provide a 100% guaranteed identification as currently each technology faces some weaknesses, especially when not adapted to the context and/or population.

The key for choosing the right one is to determine where the system will be successful and how to implement it correctly. For that purpose, the following section presents some attributes that can help clarify the different biometric modalities.

Biometrics cannot be stolen in the same manner as tokens, keys, cards or other objects used for traditional user authentication. In the same way, users can neither pass nor lend their biometric characteristics to other users as easily as they can with their cards or passwords, nor lose or forget them because of their intrinsic nature.

This is an advantage for the users as well as for the system administrators as it can save some problems and costs of system management such as, for example, the costs associated with lost, reissued or temporarily issued tokens/cards/passwords.

Biometric recognition, because of its intrinsic probabilistic nature, involves matching within a tolerance of approximation (the score) of the captured biometric against previously collected (enrolment) data. Approximate matching is required due to the variations both within and between class (i.e. individuals) in physical and behavioural attributes.

Hence biometric-related results are not as binary as the rest of IT systems [32].

However biometric characteristics are not secret. We all carry them in front of everybody.

In addition they can be either stolen from computer systems or networks that are not be properly protected or even forged [99] using dead or artificial biometric characteristics.

3.5.2 Criteria for a biometric taxonomy

As shown previously (see above) there is a large number of biometrics that can be used for authentication of individuals. Each one of them has its own strengths which depend on the expected application.

The attributes listed below, in no specific order, can help make a choice regarding what appropriate biometrics would be. Each component could be weighted or assessed individually according to the requirements of the project [43, 112].

- **Uniqueness / distinctiveness** is the level of discrimination between individuals, i.e. accepting a given person while rejecting others. The importance of this particular attribute increases with the size of the population among which people are to be distinguished.
- **Permanence** relates to the invariance or stability of the given biometric over time and context. There are a number of reasons that might alter the biometrics of a person such as age, disease, accident, clothes, skin plasticity (e.g., smiling), etc.
  A biometric should use enough features to minimise the change's effect on the system's ability to discriminate.

  This might be of less importance when processes allow for periodic update of the biometric template, i.e. re-enrolment.
- **Scalability** refers to the efficiency capability of processing of a chosen biometric in both cases of the capture step, i.e. enrolment and claim.

  This attribute is more important in large identification system contexts.
- **Convenience** relates to the disturbance degree (ease of use, speed of capture and/or treatment) of a given biometric.

  It is a factor that will be more important for disabled people (e.g., mobility or vision impaired) even though it is becoming a requirement common to all users, especially when system will be used frequently like the access control.
- **Universality** means that all (or almost) members of the population should possess the trait, otherwise it would be inconvenient and trigger difficulties for the management of security.

- **Performance**, i.e. the reliability of the chosen biometric, which means that the system's results should remain the same even if the environment might alter (e.g., lighting, temperature, etc.).
- **Vulnerability** is the system's ability to deceive fraud.
- **Privacy** is ideally protecting users by getting their permission before acquisition and strong encryption of the stored data.
- **Maintenance** is for systems using sensors requiring physical contact which potentially require more maintenance because of residue build-up, wear-and-tear...
- **Health**: Capturing biometric should be painless and harmless.
- **Quality** of the sample captured is important to ensure accurate testing of the matching potential. The required quality should be obtained easily with the chosen system.
- **Integration**: the chosen biometric should be able to come in addition to other authentication test in order to build a multi-factor authentication system.
- **Cost / benefit appreciation**: Expected benefits could be enhanced security, convenience, smaller cost of token to be replaced, suppression of human operator, etc.
- **Acceptance** by the population to be submitted to the system.

There might be some additional criteria to take into account depending on specific context. All of the above should be considered as important for the reliability and efficiency of the solution to be implemented.

3.5.3 Current limits of biometric authentication

Based on [32] and [99], it appears that a number of domains need further research especially in the following three areas:

- Human recognition systems are intrinsically probabilistic, and thus potentially fallible. The odds of an error can be drastically reduced but not totally eradicated, which should be taken care of by system designers and operators.
- Fundamental understanding of biometrics needs strengthening, since understanding the biometric traits distributions within given populations to the variation of human interaction with biometric systems.

  This could allow for defining methods enabling the acquisition of rigorous biometric measures, especially clarifying required operational and environmental conditions.

This is important particularly for biometric systems that are to be deployed at national level.

- Approaching a biometric solution from a full system perspective is a critical success factor. This involves taking complex decisions regarding definitional, technological, and operational possibilities and appreciating the technological and social contexts in which they are integrated, in order to ensure performance, effectiveness, trustworthiness, suitability and security of automated solution.

Chapter 4

Gesture Recognition

Contents

This chapter gives some background information about gesture recognition systems and an overview of the methods currently used.

The chapter presents also the devices used for Gesture Recognition biometric capture in the following of the thesis i.e. the Kinect^{TM}and the Leap Motion, as well as the comparison algorithm i.e. the Dynamic Time Warping

4.1 General background

A large part of human communication is done through gesture as shown by the works of [51] on body language, which he called proxemics. Through gesture, emotions, intentions and messages of all kinds can be exchanged without saying a word.

Gestures can be made using any part(s) of the body, and especially hands, arms, legs, fingers, or the head as well as facial expressions. A combination of several of the former is even more expressive.

In fact humans use gestures for several purposes, which can even be combined in one same gesture. They have been called: semiotic, ergodic, and epistemic [25]:

- **Semiotic gestures** are the very common gestures aimed at conveying meaningful information to others. It is culturally coloured, i.e. the same gesture in two different places and context can have very different meaning.
  The thumb-up gesture is a good example of such diversity [42], e.g, for British people it generally means “brilliant”, i.e. a congratulation or an approbation but here below are some examples of its other meanings:
  - in American sign language, it can mean “yourself” (if thrust toward another person) or the number 10 (when slightly shake left and right)
  - for scubadivers it means “going upwards/returning to surface”
  - for hitchhikers, it is a way to ask car drivers passing by for a lift
  - in Brazil it means “obrigado-thank you”
  - in ancient Rome, it seems it was a vote for “sparing the life”
  - in the middle-east, Iran, west and south-Africa it has a very bad meaning comparable to ‘giving someone the finger’
- **Ergotic gestures** relate to manipulating or interacting with the environment, such as when creating tools or artefacts, or wiping dust from a table, etc.
- **Epistemic gestures** relate to using tactile experience when discovering the environment. When touching something humans can appreciate its resistance, texture, structure, etc. Babies discover their world this way.

Any gesture of each of the above types may be emphasized using an object, e.g., a handkerchief when moving the hand in a ‘good-bye’ gesture, a brush for wiping dust, or a stick to explore the world.

4.2 Overview of the methods used for gesture recognition

Computer-based Gesture Recognition (GR) uses mathematical algorithms to capture the gesture in itself and then to understand the meaning of that movement. Developing computer-based GR allows for providing a more convenient and natural human-machine interface, i.e. interacting in a natural way not necessarily requiring the use of any particular device, such as keyboards or mice.

GR is currently developing and is now accessible to common users.

These capabilities cover the full range of the three types of gesture listed above. For example it is possible to control the mouse cursor of the computer (Ergodic), to translate sign language (Semiotic) or to analyse some sporting movement (Epistemic).

Such solutions can be used in many areas of everyday life.
For example, they help make game playing experience more genuine through increased interactivity with the virtual machine, or can analyse movements in order to identify potential mistakes and rectify them.
This can be used for high-level practitioners in dance, high level sport or other activities requiring appropriate gesture, or even to translate sign language into words, either written or spoken.

There are many ways to recognise and track a person's movements, using different kinds of sensors. They can be classed into two main categories, hand-held device-based ones and hands-free ones.

- Hand-held controlling devices of some kind are the most common way of controlling a computer.
  - The mouse of a computer is the one used by most users in the world, whether internal or external, wire-linked or wireless.
  - This category includes all other instrumented devices such as gloves, body suits, and marker-based optical tracking which require holding something that helps to track movements.
  - Motion sensing devices that embed accelerometers and gyroscopes (to measure acceleration along the three axes x, y and z over time) with optical sensors (for pointing) like the Nintendo's Wii remote control for the Wii console, are a different way of controlling a computer that has appeared and developed some years ago.

    This answers the users' expectations which are very clearly expressed when using the terminology “wirefree” instead of “wireless”.

PDAs are already equipped with such technology and soon they might be embedded in watches [65].

- The vision-based way is a hands-free approach which is literally viewing the user's gesture(s) through sensors such as webcams. This category is also named depth camera.

  The vision-based approach can rely on several kinds of sensor technologies of which are listed below. It can thus be based on:

  - A single camera is the most accessible. It is a basic camera that can recognise gestures. Such basic devices are limited to capture and recognise only 2D (two-dimensional) gestures.
  - A stereo camera relates to two cameras positioned next to each other.

    Such equipment allows for capturing and working out the depth of the situation by corresponding points in the two images. This is the way human eyes determine the depth of what we are seeing.
  - A camera array is the extended version of the stereo camera as it is using multiple cameras to capture an action from many angles at the same time.

    All the cameras have to work as one single system for properly estimating 3D (three-dimensional) scene geometry from the dense imagery (light fields) captured by the array to construct multi-perspective panoramas.
  - A time-of-flight camera is a range imaging camera which takes its name from the fact that such systems are built using methods comparable to radars or LIDAR (Light Detection And Ranging which are used to measure distance and 'illuminate' a target), which take into account the light travel time (based on the speed of light) to resolve distance.
  - A structured-light 3D scanner determines the 3D structure of a scene based on the captured distortion of the projected pattern (infrared light).

    The Figure 4.1 shows how such a device works.

    The Kinect^{TM} sensor is good example of this kind of device.

Choosing a sensing technology requires an analysis of several factors such as resolution, latency, range of motion, user comfort, and cost [83]. Although nobody can deny the importance of cost in such an approach, the high-paced continuously decreasing trend of technology market prices that has been seen for more than a decade, makes it very relevant.
As a matter of fact some devices that were considered very expensive less than 10 years ago are now quite cheap even in their up-to-date version.

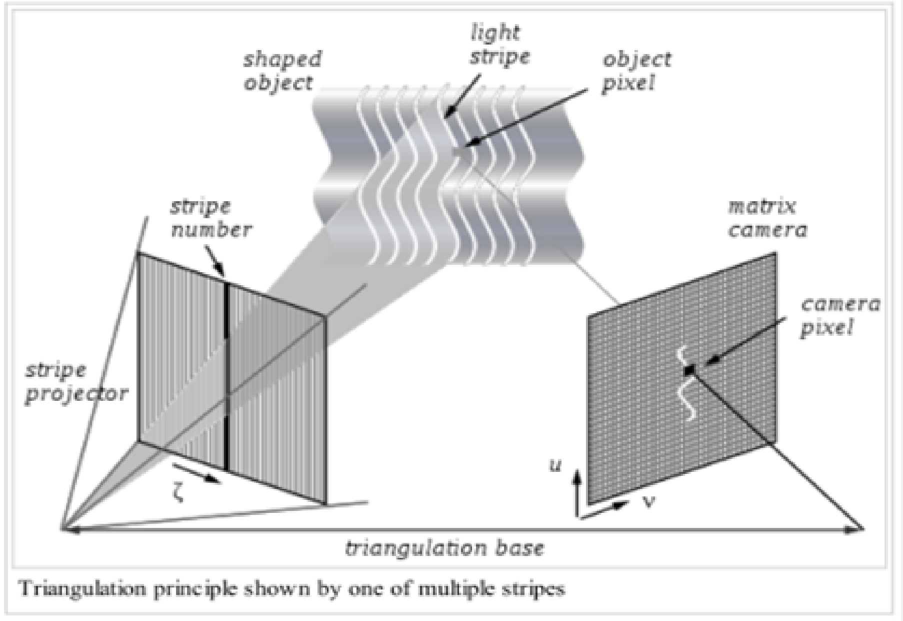


Figure 4.1: *Working principle of a structured-light 3D scanner (source [5])*

One can also understand from the market evolution that the user comfort dimension has become a dominant one. With this in mind, hand-held devices are, if not exactly doomed, rather cumbersome as even if very small they hinder the ease and naturalness of the user's interaction with the computer. However most vision-based techniques (except the last one mentioned above), while overcoming this, still need to overcome some problems, in particular those related to occlusion of parts of the user's body say [83]. And this statement might very soon be outdated.

4.3 The gesture recognition system used

Accurate GR depends on factors similar to those mentioned above for biometrics in general, i.e.:

- The sensor(s): The quality of the input which, in other words, relates to precision of the captured points, the camera location, lighting, potential shadows, etc.
  This is related to the capture step of biometric authentication.
- The feature extraction: The appropriate selection of captured features, i.e. choosing and making sure to acquire the key points adapted to the gesture.
  This is equivalent to the use of biometric sensors.

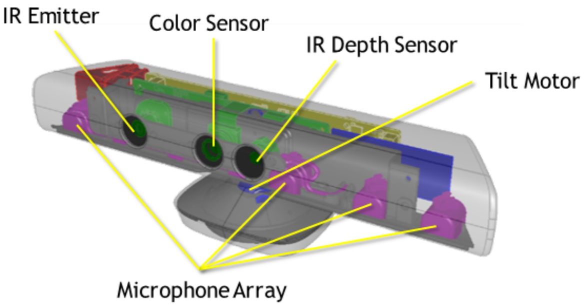


Figure 4.2: *The Kinect^{TM}(source [81])*

- The analysis: The choice of the detection algorithm for an appropriate interpretation.

The next section provides some more information on each of these.

Comparisons with biometric authentication systems can be made as gestures have physical characteristics and often behavioural ones too.

There are two main ways to achieve GR [83]:

- The model-based approach which uses the 3D coordinates of some key body parts to get several important data points such as (but not only) the joints (knee, ankle, elbow, shoulder). Two sub-kinds need to be distinguished, i.e.:
  - volumetric models which are preferably used in offline algorithms,
  - skeletal models used for real time analysis, as only key parameters have to be taken into account.
- The appearance-based approach which uses image sequences or features derived from these as gesture templates and thus consists in reducing the error between the input image and the closest model instance.

4.4 The sensors

4.4.1 The Kinect^{TM}

The Kinect^{TM}, shown in Figure 4.2, is a good example of a structured-light 3D scanner.
It was originally developed as a completely hands-free control for a Microsoft gaming console, the Xbox 360, in order to offer a serious alternative to Nintendo's Wii Remote control and Sony's PlayStation Move-Eye motion controllers.

The Kinect$^{\text{TM}}$ "understands" gestures using software created by the "Rare" company (Microsoft Game Studios group), coupled with the particular range camera technology developed by the Israeli company "PrimeSense" [78].
It can accurately track and capture under any ambient lighting, in 30 frames/sec video [80], a skeleton composed of 20 points [100] which represent the 3D position of the head, neck, spine, centre hip, as well as the left and right side joints: hand, wrist, elbow, shoulder, hip, knee, ankle and foot.
It uses an infrared projector [79] and camera fitted with a special microchip, known as Light Coding, and uses a variant of image-based 3D reconstruction to provide a 3D scanner system.

4.4.2 The Leap Motion

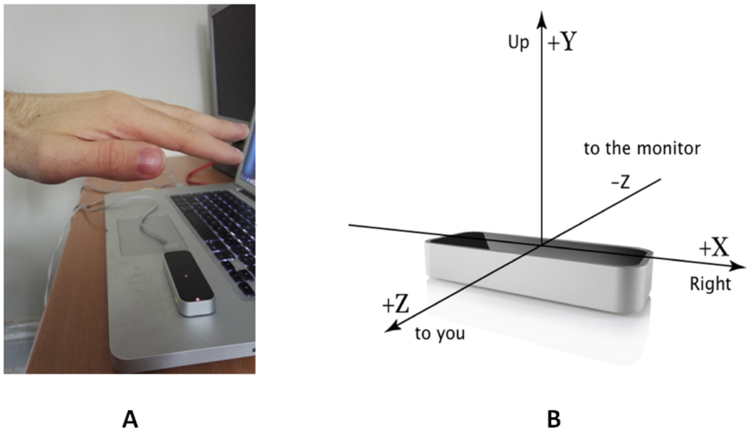


Figure 4.3: *A: User using the Leap Motion (source: Benoit Ducray) B: Leap Motion orientation (source [3]).*

The Leap Motion Controller is a device launched on July 2013 by Leap Motion [3]. Its aim is to track hands and recognise gestures in order to give the user a new interface to communicate with machines (i.e. computer, smart phone etc.).
It is composed of three IR emitters and two IR cameras. Thus it is an optical tracking system based on stereo vision, because the LEDs (Light Emitting Diodes) are generating patternless IR light.
The manufacturer claims that the sensor has an accuracy in fingertip position detection of approximately 0.01 mm. The accuracy of the Leap Motion has been discussed and it was found that the theoretical accuracy of 0.01 mm could not be achieved under real conditions; an overall average accuracy of 0.7 mm was achieved [114].
The Leap Motion device gives the position of an element in three dimensions by using coordinates (x, y, z). See Figure 4.3.

4.5 Feature extraction

Feature extraction is responsible for isolating the appropriate data amongst the amount of information captured by the sensor signal, i.e. reducing the number of input data to the amount required in order to perform a desired task. It is in particularly important to identify the beginning and the end of a gesture in order to keep only what is for the analysis part.
This is the most complex part as the same gesture can present important variations not only between two different persons but even when accomplished several times by one single person.

Other background data is captured by the sensor that brings no additional information on the nature and meaning of the gesture and thus needs no treatment during the analysis phase. This background data can also be eliminated.
Such an elimination process normally requires either expert knowledge of the data to analyse or applying data-driven dimension reduction techniques such as principal components analysis (PCA), independent component analysis (ICA), or linear discriminant analysis (LDA). Often a combination of the three approaches can be used.

4.6 Analysis

They are several algorithms to do GR: Dynamic Time Warping (DTW), Neural Network or Hidden Markov Model. We have mainly focused on the DTW algorithm, because it requires little or no learning period for the machine.
Other classification algorithms such as Neural Network or Hidden Markov Model need several examples from the user in order to get an accurate authentication rate.

4.6.1 Dynamic time warping

Dynamic Time Warping is a type of algorithm which uses a temporal structure technique. It is used to find an optimal alignment between two time-bound sequences, independently of the variation of time or speed between both sequences.

Originally, this algorithm was used in speech recognition [113] and its use has been enlarged to all domains in which data can be modelled in a linear representation e.g., computer animation, video, audio and graphics.
One of the applications is for GR to extend the interface between human and machine [38]. The interested reader is referred to other works that have used this method [46, 106, 69].

This capability of finding an alignment for two sequences which are comparable but not aligned is very important when comparing gesture patterns.
The speed at which a given gesture is performed can be highly variable from one occurrence to another, even when the same individual is performing them in a sequence.

In practice, the principle of DTW is to define a warping path with the minimal cost. This cost is given by the cost function (or distance function) which is the distance (or the error) between the two sequences, as shown in Figure 4.4.

In other words, the DTW algorithm gives the alignment between the curve given by the reference model data, and the curve given by the user's captured data. DTW is reviewed in [63] and can be summarised as follows:

In order to use DTW algorithm to align two sequences A and B, where:

- $A = (a_1, a_2, ..., a_N)$ of length $N \in \mathbb{N}$ (i.e. a positive integer)
- $B = (b_1, b_2, ..., b_M)$ of length $M \in \mathbb{N}$,

We construct an N-by-M matrix where the (i^{th}, j^{th}) element of the matrix contains the distance $d\ (x_i, y_j)$ between the two points x_i and y_j, using a distance function, generally the Euclidean distance, $d(x_i, y_j) = (x_i - y_j)^2$.

Each element (i, j) of the matrix corresponds to a hypothetical alignment between the points x_i and y_j.

From this matrix we can determine a warping path W where the k^{th} element of W is defined as $w_k = (i, j)_k$ and we thus have:

$$\begin{aligned} W =& w_1, w_2, \dots, w_k, \dots, w_K \\ max(m, n) \leq& K < m + n - 1 \end{aligned} \tag{4.1}$$

The warping path is typically subject to constraints on boundary conditions, continuity and monotonicity with the following constraints:

- **Boundary conditions**: $w_1 = (1, 1)$ and $w_k = (m, n)$, the warping path must start and finish in diagonally opposite corner cells of the matrix.
- **Continuity**: Given $w_k = (x, y)$ then $w_{k-1} = (x', y')$ where $x - x' \leq 1$ and $y - y' \leq 1$. Allowable steps in the warping path are restricted to adjacent cells (including diagonally adjacent cells).

- **Monotonicity**: Given $w_k = (x, y)$ then $w_{k-1} = (x', y')$ where $x - x' \geq 0$ and $y - y' \geq 0$. The points in W are forced to be monotonically spaced in time.

We are interested only in the path which minimises the warping cost:

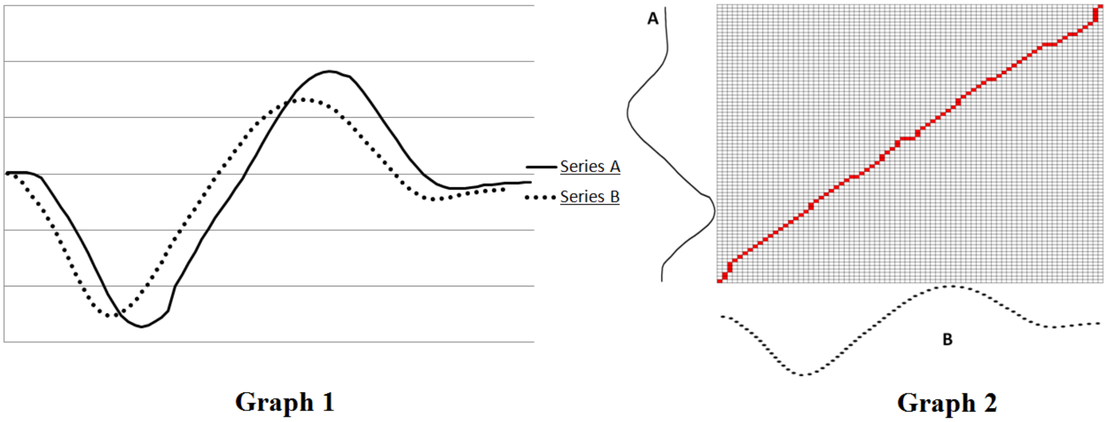
$$DTW(AB) = min(\sqrt{\sum_{k=1}^{K} w_k}) \tag{4.2}$$

We can find this path using dynamic programming to evaluate the following recurrence which defines the cumulative distance $\gamma(i, j)$ as the distance $d(i, j)$ found in the current cell and the minimum of the cumulative distances of the adjacent elements:

$$\gamma\ (m; n) = d(m; n) + min(\gamma\ (m-1; n-1); \gamma(m-1; n); \gamma(m; n-1)) \tag{4.3}$$

Where: $\gamma\ (m; n)$ is an $(M + 1) \times (N + 1)$ matrix; $\gamma\ (0; n)$ and $\gamma\ (m; 0)$ are initialized with a large number representing infinity or zero, depending on the application; $\gamma\ (0; 0)$ with zero; $d(m; n)$ is the cost function.

The cost of the minimal path between both sequences is contained in $\gamma(M; N)$.



Graph 1 **Graph 2**

Figure 4.4: *Illustration of the principle of the Dynamic Time Warping algorithm*
Graph 1: Two time series (A and B) - Graph 2: The warping path between A and B obtained using the DTW algorithm (source: Benoit Ducray)

Part II

Contributions

Chapter 5

Dynamic versus Fixed Biometrics Security

Contents

Biometric data can be used as “something you are” in authentication systems, but if a biometric is compromised by a malicious entity, the genuine user can no longer use it because it cannot be easily changed.
What we defined as ‘Dynamic biometrics’ may offer a practical alternative, as they capture both an inherent factor along with a changeable knowledge factor in a single step.
This chapter investigates dynamic biometrics and whether they offer useful security authentication properties compared to conventional biometrics.

Security characteristics of examples from three classes of dynamic biometrics are compared to a selection of common physiological (“fixed”) biometrics, leading to the conclusion that in addition to providing one-step, two factor authentication, dynamic biometry may provide privacy benefits in some circumstances.

5.1 Problem statement

Biometrics are often used in authentication solutions to provide “something you are”. However, attackers may seek to compromise biometric authentication; possible attacks include compromise of stored biometric data, or copying or faking biometrics to fool data capture sensors. The compromise of a fixed user biometric is a fundamental disadvantage of this type of authentication, so the use of dynamic (changeable) biometrics may provide a practical alternative.

The availability of new types of sensors such as depth cameras, brainwave sensing headsets, etc., has generated research interest into what we called ‘dynamic biometrics’, as these sensors can be used to capture inherent factors (physical/behavioural) simultaneously with a knowledge factor; e.g., Gesture Recognition (GR) (e.g. [71, 15, 40]).

In the same way that the knowledge factor is easily changeable, a dynamic biometric can be changed yet still retain the advantages of biometric input. It will also provide a means for one step two-factor authentication [27], where only one action is required to present two authenticating factors to a verifier.

This chapter investigates whether dynamic biometrics can provide security authentication and compares them to fixed (static) biometrics. Several examples of dynamic biometrics are presented. The security characteristics of several fixed or dynamic biometrics are determined based on criteria devised by Bonneau et al. [23]. We chose to take Bonneau et al. as a reference for the security criteria as they have done a useful work to compare various web authentication schemes. In this document, we will focus on the criteria which are relevant for biometrics schemes but future research can easily associate the analysis done here with Bonneau et al. work to extend the comparison.

The three fixed (static) biometrics included in the analysis are:

- *Fingerprints* as they are a well proven and widely used biometric [53];
- *Face Recognition*, which is also a widely used and accepted biometric [53];
- *Retina* as it is seen as a highly reliable and accurate identifier [30].

We chose to use this three (static) biometrics as they are among well proven and widely used for the Fingerprints and Face Recognition [53], while Retina recognition is one the of the most highly reliable and accurate biometric [30].

This chapter is structured as follows: Section 5.2 presents the limits of fixed biometrics and the solutions proposed. Section 5.3 explains the background about dynamic biometrics and defines its different categories. Section 5.4 shows evaluation criteria and the security assessment of each biometric, and results are analysed in Section 5.5.

5.2 The need for the concept of ‘Dynamic Biometrics’

5.2.1 The limit of biometrics

Compared to other authentication factors, biometrics may seem without fault, as they are not easily stolen, forgotten, or lost. The main limitation of biometrics is that they are not secret − they are public data − hence biometric data can be captured and replayed.

For example, when considering the most widely used biometric, fingerprints [53], it is relatively easy to obtain fake samples because latent fingerprints are left on every surface we touch.

Several people have had their biometrics stolen: for example, the defence minister Ursula von der Leyen had her thumbprint replicated by a member of the Chaos Computer Club [64]; or the hack of a US government where 5.6 million federal employees had their fingerprints stolen [50].

Alarm bells have been raised by a team at Japan’s National Institute of Informatics about the popular two-fingered pose as the team was able to copy fingerprints based on photos taken by a digital camera three metres away from the subject [6].

5.2.2 Partial solutions to such limitations

Some works propose the use of cancelable biometrics [91, 12, 92] to reach the revocability of biometric. The main point of this technique is to store a transformed biometry using a one-way function and the transformed biometric and the transformation are both retained.

This construct preserves privacy since it is not possible (or computationally very hard) to recover the original biometric template using such a transformed version.

If a biometric is compromised, it can be simply re-enrolled using another transformation function, thus providing revocability [91].

But this is offering protection of the biometric main template on the storage point, but provides no protection at all in case of the original biometric data being stolen.

Another solution to protect against fake biometrics is to detect the aliveness of the sample. The method used to determine whether the biometric is alive or not depends on the biometric and system used. Indeed, contact and contactless systems do not have the same ability to determine aliveness.

There are two major ways to do this, either with the hardware, which needs to be performed at the point acquisition, or software which would be performed at the processing stage.

For example, when using fingerprints, we would have:

- From a hardware point of view, the countermeasures possible are: skin temperature, optical properties of the skin, pulse oximetry, blood pressure, electric resistance of the skin [74, 98].
- From a software point of view, the countermeasures possible are: Local Binary Patterns [87], Pores detection [76], Power spectrum [31], Wavelet energy signature [85], Ridges wavelet [104], Valleys wavelet [105], Curvelet energy signature, Curvelet co-occurrence signature [84].

If any weakness is found in one of the hardware countermeasure, it would be very expensive to change the full system.

According to [102], even if the authors claim a very high performance score, the performance of their methods depends on knowledge of the fake biometric fabrication techniques and materials during the development of the method.

5.3 Dynamic biometrics

5.3.1 Definition of dynamic biometrics

Our definition of dynamic biometrics is shown below (other papers have used slightly different definitions e.g., [101]):

A biometric is dynamic when *physical/behavioural (inherent) biometric information is captured together with a knowledge factor from a user, such that it can be used as the basis of a one-step two factor authentication.*

5.3.2 The three classes of dynamic biometrics

We introduce the following three dynamic biometric classes: text based, gesture based, and thought based.

Text based

Keystroke, Speaker Recognition and touch screen patterns[1] on smart devices are good examples of this class [27], provided that the text/pattern used has been chosen by the user. Here, there is both biometric information (either the keystroke, sound emission or touch screen speed/style/pressure) and something the user has to know and can change easily, i.e. the text.

Thought based

Authentication based on brainwave signals is now a realistic possibility. Several works have proposed the idea of a “passthought” and have shown than it is possible to authenticate a person via a specific thought [107, 61, 28]. Here the knowledge factor is the particular thought and the inherence factor is the uniqueness of the brain’s wave emissions [107].

Gesture based

This class can be divided into Gesture and Signature categories.

Gesture: There are several ways to capture a gesture: for example, by using a depth camera (described later in this paper); or by using an electromyograph to capture electric impulses in the muscle [54]. The knowledge factor is the gesture itself.

Signature: This refers to the capture of the direction, stroke, pressure, and shape of a signature, through touch sensitive technologies (such as PDAs or tablets). This only matches our definition of a dynamic biometric when the handwritten text can vary.

5.3.3 Advantages and disadvantages of dynamic biometrics

Dynamic biometrics have several advantages; they are easily changeable due to the knowledge factor, and they allow one-step two-factor authentication[2]. However, some categories of the dynamic biometrics family only use weak physical biometrics, such as gestures based on

[1] Touchstroke dynamics is a behavioural biometric based on the style and rhythm that someone uses to interact with a touchscreen-equipped smartphone. This authentication method is analysed in [37] and [62], and enhanced in [11] e.g., by proposing how to handle typos.

[2] However the reverse is not true, all one-step two-factor authentications are not dynamic biometrics e.g. the Bionym wristband allows authentication via the cardiac rhythm it records plus ownership of the wristband itself [2].

upper body geometry (shoulder length, arms length) which may be a disadvantage in some situations. Furthermore, any behavioural elements in the biometric may be observed/copied and knowledge factors may be forgotten.

5.4 Fixed vs dynamic: an evaluation framework

The evaluation of dynamic biometrics that follows is based on security criteria outlined in the work of Bonneau et al. [23]. Out of the 25 criteria described in Bonneau et al. work, we chose to focus on the seven criteria relevant for biometrics along with additional security criteria that are particularly relevant to dynamic biometry. We have chosen to include in this comparison one example of each dynamic biometric class.

We chose *Speaker Recognition* rather than *Keystroke Recognition* or touch screen because it is easily deployable over existing communications infrastructure (the telephone system) [73] and stable over various devices (Keystroke and touch screen are not)[3].

In the end, we selected *Gesture Recognition* and *Passthought* as they are both relatively new research areas.

Note that we expand **the definition of 'attacker'** that here refers to an individual who attempts to obtain biometric information by any method (not just copying) in order to successfully authenticate in the stead of the genuine user.

5.4.1 Evaluation criteria of the biometric types

- Resilient-to-Physical-Observation:
  if an attacker is present when the genuine user is authenticating, they should not be able to capture any useful information.
  We rate a biometric as High if no information can be captured, Medium if some information could be captured and Low if almost all the information can be captured.

- Resilient-to-Targeted-Impersonation:
  if an attacker has investigated background information about the genuine user, they should not be able to use this successfully in authentication.
  We rate a biometric as High if no information can be captured, Medium if some information could be captured and Low if almost all the information can be captured.

- Resilient-to-Unthrottled-Guessing:
  the system should not permit an attacker to be authenticated if they are allowed unlimited

[3] However, touch screen biometrics on smart phone devices exhibit many of the same characteristics as speaker recognition

tries.
We rate a biometric as High if the attacker would need more than 2^{20} attempts, Medium if they need more than 2^{10} attempts and Low if they need less than 2^{10} attempts.

- Resilient-to-Theft:
  if the system uses a physical object for authentication (i.e. reader, keyboard, etc.), this object should not give any information to an attacker if they get access to it.
  We rate a biometric as High if it does not need any physical object or the object does not keep any information, Medium if the biometric might require an object that an attacker could get information from, and Low if the biometric always needs an object that an attacker could potentially get information from.

- Requiring-Explicit-Consent:
  Here we rate a biometric High if it needs the full consent of the user to start an authentication process, Medium if the biometric can be used to authenticate without the consent of the user only by using subterfuge, and Low if the biometric can be used to authenticate without the consent of the user.

- Unlinkable:
  For privacy, it should not be possible for colluding verifiers to determine if the same user is authenticating to both their systems with the same secret.
  We rate a biometric as High if there is no linkability, Medium if it is linkable in some circumstances and Low if it is totally linkable.

- One Step Two Factors:
  Does the biometric combine two factors in one step? (yes or no)

- Changeable:
  Can the biometric be changed and reused for authentication in the case of compromise? (yes or no)

- FAR / TAR:
  False Acceptance Rate (FAR): an attacker successfully authenticates - True Acceptance Rate (TAR): a genuine user successfully authenticates.

5.4.2 Evaluation of biometrics

The following analysis has been summarised in Table 5.1.

Table 5.1: *Comparison of biometrics (source Benoit Ducray).*

	Physical Biometrics			Dynamic Biometrics		
	Fingerprint Recognition	Face Recognition	Retina Recognition	Speaker Recognition	Passthought	Gesture Recognition
Resilient to Physical Observation	High	Low	High	Low	High	Medium
Resilient to Targeted Impersonation	Low	Low	High	Medium	Medium	Medium
Resilient to Unthrottled Guessing	Low	Low	High	Low	Low	Low
Resilient to Theft	Medium	High	High	High	High	High
Requiring Explicit Consent	Medium	Low	High	Medium	High	High
Unlinkable	Low	Low	Low	Medium	Medium	Medium
One Step Two Factor	No	No	No	Yes	Yes	Yes
Changeable	No	No	No	Yes	Yes	Yes
False Acceptance Rate	0.2%	0.1%	10^{-7}%	2%-5%	2%	0%-3.4%
True Accept Rate	99.8%	90%	99.99%	80-90%	98%	81%-99%

- Resilient-to-Physical-Observation:
  We rated Fingerprint and Retina as High, along with Passthought: for Passthought an observer cannot capture what the user is thinking as there is no device yet that can capture brain waves at a distance.
  GR is rated Medium, as although an observer can see/observe/record a gesture, they cannot use a recording directly to get authenticated. Face and Speaker Recognition are rated Low.

- Resilient-to-Targeted-Impersonation:
  Retina is rated High as it is difficult for an attacker to find out the blood vessel pattern.
  Speaker Recognition, Passthought and Gesture are Medium because the attacker may be

able to discover information relevant to the specific thought and gesture but that would not be enough to perform an attack.
Fingerprint and Face Recognition are rated Low as it would be easy for an attacker to find a picture [109] and a latent fingerprint [75] to impersonate a user.

- Resilient-to-Unthrottled-Guessing:
Here we will use FAR data with the formula given in [86] to calculate the keyspace which is $1/FAR = keyspace$.
We base our ratings on the following calculated keyspaces: Fingerprint $2^{8.97}$, Face Recognition $2^{6.7}$, Retina $2^{29.89}$, Speaker Recognition goes from $2^{4.35}$ to $2^{5.65}$, Passthought $2^{5.65}$. Gesture[4] ranges from $2^{4.9}$ to $2^{6.65}$.
We can see that, according to the above criteria, all here listed biometrics are Low except Retina which is High.

- Resilient-to-Theft:
All the biometrics analysed here do not required any contact with an object or leave any information on it, with the exception of Fingerprint. This leaves some latent prints on the reader, which provides a way to attack it [75].

- Requiring-Explicit-Consent:
Retina scanning requires the user to look into an eyepiece and focus on a specific spot [60]: we rate this High.
Similarly, a High rating was given for Passthought and Gesture as it would be difficult for a attacker to authenticate without user consent.
Fingerprint is rated Medium as an attacker could trick a genuine user into touching a reader to initiate an authentication.
Speaker Recognition is rated Medium because an attacker could use a hidden microphone to authenticate as a genuine user without their consent.
Face Recognition is rated Low as an attacker could authenticate using an easily obtained photo of the user taken without their consent [109].

- Unlinkable:
By definition, biometrics are related to a particular user, so all 'fixed' biometrics are rated Low. However, some dynamic biometrics use weak inherent biometric factors: for example GR may involve some body measurements such as arm length or shoulder width [40] which are not sufficient for a unique identification of a user. This make them better for privacy. Also the inclusion of a knowledge factor in dynamic biometrics means

[4] An FAR of 0% was found in [40], i.e. a very high keyspace, but we feel that this value needs to be confirmed by further experiments.

that the same inherent factor can be used with different secrets at different verifiers. Consequently, all dynamic biometrics are rated Medium.

- One Step Two Factors:
  None of the physical biometrics can be used in a One Step Two Factors authentication, but by definition any dynamic biometric can.
- Changeable:
  Physical biometrics cannot be changed at the wish of the user. With dynamic biometrics the knowledge factor can be easily changed.
- FAR and TAR:
  For Fingerprint, Face, Speaker Recogniton we based this section on [60].
  For Retina, the error rate is 0.0000001% [30] so we can assume than the FAR is the same and the TAR is 99.9999999%.
  Passthought [61] found a FAR of 2% and a TAR of 98%. For GR we used the range of values from Table 6.3.

5.5 Security of dynamic vs fixed biometrics

The data shown in Table 5.1 highlights some important issues.

There have always been concerns about privacy and linkability of biometrics, and that once compromised, a biometric credential becomes unusable by the genuine user.

These concerns are addressed by dynamic biometrics, and it can be seen from the table that this new family of biometrics outperforms some traditional biometrics in a number of respects. For example, our rating of Face Recognition is equal to or lower than all the dynamic biometrics assessed, for all security criteria identified.

We can also see that Passthought could rival Retina Recognition in terms of security, being ranked lower in only two criteria, *resilient-to-targeted-impersonation* and *resilient-to-unthrottled-guessing*.

All the dynamic biometry categories were ranked Medium in the *resilient-to-targeted-impersonation* criterion, better than Fingerprint and Face Recognition.

Naturally, not all biometrics are suitable for all authentication situations: conventional biometric techniques are typically used for applications with higher security requirements than dynamic biometrics.

For example, as Gesture authentication is vulnerable to ‘shoulder-surfing’ (copying) attacks it would not be suitable for use in busy public environments without the use of some kind of voting booth concealing the person executing the gesture, but would be a plausible option for video games.

Passthought currently requires fairly intrusive use of hardware so may not be a good option for day-to-day use.

Dynamic biometrics are by definition capable of providing One-step Two Factor Authentication, and the use of a secret knowledge factor brings some privacy benefits in comparison to ‘fixed’ biometrics: additionally the use of weak inherent biometric data in gestures will also improve Unlinkability.

5.6 Summary

A major security issue with ‘fixed’ biometrics occurs if biometric data is compromised so the use of a changeable, dynamic biometric may provide a practical alternative. This thesis investigated how the security of dynamic biometrics compares to conventional biometrics. Several examples of dynamic biometry were presented.

Evaluation criteria devised by Bonneau et al. [23] were then used as a basis to assess the security of several fixed/dynamic biometrics.
The inclusion of a knowledge factor in a dynamic biometric brings some privacy benefits in comparison to ‘fixed’ biometrics, in addition to making the biometric changeable.

Unlinkability improves a) because the same physical characteristic can be used at different verifiers with different secret knowledge, and b) by using weak inherent biometric data in some dynamic biometrics (e.g. in GR).

Chapter 6

Feasibility of Authentication based on Gesture Recognition

Contents

In this chapter, we take an example of dynamic biometric, the Gesture Recognition (GR). We study the ability of a system based on GR to authenticate genuine user and reject an impostor.

For this we identify what are the possible attacks against a GR: "Brute Force attack", "Dictionary attack", "Storage Leakage" and "Shoulder Surfing attack". We set up two experiments which test the GR system against "Brute Force attack" and "Shoulder Surfing attack".

For these experiments, we used two different sensors: the Kinect^{TM} design to recognise gesture made by the upper body and the Leap Motion, designed to recognise a movements made by a hand.

We compared the results with others studies. Our results align with those found by others work on this domain.

6.1 Problem statement

Since the commercialization of the Kinect$^{\text{TM}}$ in 2010 several devices propose to get depth information of a scene either by being a depth camera such as the Leap Motion or by embedding a depth camera such as the Lenovo Phab 2, which provides depth and motion tracking sensors. This proliferation of depth camera will last for several years [70].

This technology may be used in several applications, including measuring distances and geolocation. It would be interesting to see if this technology, as it proliferates, could be used for authentication. Gestures are not usually aimed at high security applications, but as convenient alternatives to simple PIN or password entry.

However, depending on the method and precision of capture, gestures can include some biometric related characteristics as well as the "something you know", making them more like two-factor authentication inputs.
This type of authentication has the added advantage that in the case of compromise, the gesture can be changed yet still retain the advantages of biometric input. However, one of the obvious weaknesses is that a gesture has to be performed in plain sight, so even an unskilled observer can attempt to copy it, to attack the authentication system.

In this chapter, we are proposing to study the feasibility of using Gesture Recognition (GR) for authenticating a user. In order to investigate this question, we have set up two experiences:

- One using the Kinect$^{\text{TM}}$ design to recognise gesture made by the upper body (i.e. both arms movement). Ten volunteers provided ten reference gestures of their own design. These gestures were attacked by 28 people who first tried to guess the gestures and, after watching them, tried to mimic them.

- The second experiment was a proof of concept using the Leap Motion, designed to recognise a gesture made by a hand. This aimed to highlight the difficulty of mimicking a gesture. A group of ten volunteers was asked to reproduce ten predetermined gestures ten times.

  We also tested GR using just the palm centre data points. This effectively removed all hand geometry information, resulting in similar data to that which could be obtained from an accelerometer sensor.

This chapter is structured as follows: Section 6.2 gives the definition of what we call a "gesture" and how the biometric information affect the authentication. Section 6.3 presents different ways an attack can be performed on a gesture. Section 6.4 presents our proposed system, and the actual experiment for both experiences. Section 6.5 presents the results. Section 6.6 compares our work to what has been done by others on GR. A comparison with fingerprint is also done on two security abilities: the ability to correctly recognize a genuine user, and the ability to reject an individual trying to mimic the biometric.

6.2 Gesture recognition

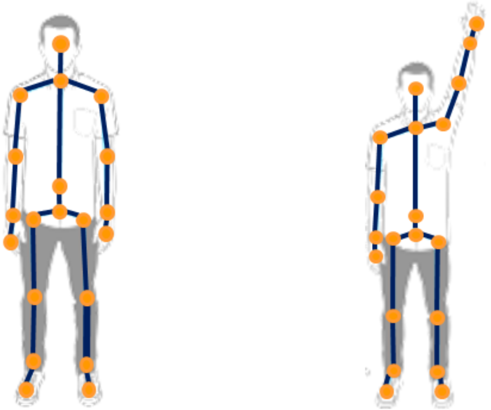


Figure 6.1: *Kinect^{TM}'s Skeleton (adapted from [79])*

When we refer to a "gesture" we mean a set of frames and tracking points produced by our gesture capture devices. These elements are organised such that every frame contains the same number of tracking points.

- Frames: These represent the length of a gesture in time. The frequency generation of a new frame depends on the sensor; the Kinect^{TM} software generates 30 frames per second [81], the Leap Motion software can produce frames up to a rate exceeding 100 frames per second [4].

- Tracking points: These represent the features of the gesture. Movement is usually tracked in three dimensions, which means that for each point we have information on the horizontal, vertical and depth (x, y, z). Depending on the sensor used, there may be support for point tracking, for example the Kinect^{TM} is able to track points from the head to toe including the head, hands, hips, feet, etc. [79] (see Figure 6.1), but it is also possible to use the raw image from the sensor to track other points. The Leap Motion device, which tracks and records hand movements, could use raw data to track more information about the hands, such as finger thickness.

Table 6.1: *Limbs' length in centimetre (data extracted from [49])*

	Men		Women	
	Average	Standard deviation	Average	Standard deviation
Lower Arm	26.99	1.57	24.34	1.55
Shoulder to Elbow	36.90	1.79	33.58	1.74
Hand Length	19.38	0.98	18.05	0.97
Hand Breadth	9.04	0.42	7.94	0.38

These tracking points are the elements which provide the biometric information through the limbs' length (either arms' length or the fingers' length).

On its own, the length of a limb does not have a variation which can be used to accurately discriminate one individual from another. For example, the average arm's length for a man is 77.68cm with a standard deviation of 3.80cm [49] (Table 6.1 shows the average and the standard deviation of limbs' length we used).

But for GR, when using an algorithm such as Dynamic Time Warping (DTW), which accumulates the error, any limb's length difference will be conspicuous.

For example, if the difference between two limbs is 1cm and the gesture lasts 30 frames (1 second with a 30 frames per second sensor), even if the gesture is reproduced perfectly, at the end there will be an error of 30cm, and thus discriminate one individual from the other.

6.3 Attacks against gestures

There are several ways to attack a system on GR. There are four different kinds of attack presented in [29]: "Brute Force attack", "Dictionary attack", "Storage Leakage" and "Shoulder Surfing attack".

6.3.1 Brute force attack

A "Brute Force attack" consists of repeatedly trying movement in order to find the right gesture. An equivalent to a Brute Force attack on text-based passwords would be trying to guess the password.

6.3.2 Dictionary attack

A "Dictionary attack" similar to a "Brute Force attack" but the attempted movements are coming from a set of more likely possibilities (such as datasets from user studies).

6.3.3 Storage leakage

Storing a template has always been a challenge for biometric authentication systems. One the thief has stolen the template, they still need to know both the recognizers structure and how to translate the stored values into gesture actions. Text-based passwords usually only store the cryptographic hash of the passwords but two given gestures are never exactly the same, so that makes comparing their hashes effectively impossible.

6.3.4 Shoulder surfing attack

"Shoulder Surfing" can also be called "Copying Attack", or "Mimic Attack". The principle of this attack is to observe and reproduce the gesture.

We can divide this attack in to two categorizes, Human and Mechanical.

- **Human**: For this category, a human reproduces the gesture in the hope that the gesture will be reproduced well enough to compensate the biometric geometric difference. To do so we can decompose the attack into three steps: the observation, the learning process, the reproduction.

    1. The attacker has to observe the gesture of the genuine user. This may be seen or recorded. The more points of view the attacker has, the more accurate their gesture reproduction can be. This can be achieved either by having some collaborator who will observe or record the gesture from a different location, or if they have the opportunity of changing their observation point every time the genuine user is doing the gesture. One of the main difficulties of this portion of the attack is to not arouse suspicion of the genuine user or any nearby guards.
    2. The learning process is not the most difficult part but there are several errors to avoid. The attacker or the collaborator who is learning the gesture should be the one who has the closest biometric characteristics to the genuine user.
       When the gesture is reproduced, it should be well sided (sagittal section) − this is difficult because we noticed during our experiments that people tend to reproduce a gesture as if has been done by a mirror.
       The learner has to pay attention to any little movements made by the genuine user such as shoulder movement, hip rotation, etc.
    3. The reproduction step occurs when the attacker is in front of the sensor and has to reproduce the gesture they have learned. In order to have more chance of success, the attacker may use spoofing in order to reduce the error introduced by the difference of geometry.

- **Mechanical**: For this category, a mechanical device (i.e. a robot) reproduces the gesture. We can decompose this attack into four steps: capturing, reconstituting, faking the biometric geometry, and reproducing.

    1. Capturing aims to record the gesture in three dimensions. Because it needs to be recorded in 3D, a suitable recorder needs to be used. One of the main challenges of this portion of the attack is to successfully place the recording device where it will be able to record the full gesture without arousing any suspicion. Several devices may be used. Subsequently the different videos may be merged in order to create the full gesture scene.
    2. Reconstituting consists of numerical extraction of the different movements of the features (i.e. the joints used for the GR), composing the gesture, and translating and rotating them in the way it must be seen by the genuine sensor, and making them understandable to the robot.
    3. Faking the biometric geometry: the main advantage of using a robot is that it does not have any specific biometric geometry but it needs to adopt the biometry of the genuine user. The hardest part of this step is to get the right mensuration of the genuine user. Once it has been done, a basic articulated mannequin can be used to perform the gesture.
    4. Reproducing: Once the robot has good geometry and can reproduce the gesture, it needs to set in front of the sensor in order to execute the attack on the system.

The Mechanical attack is a lot more accurate than a Human attack but has several drawbacks: the cost of a robot reproducing a human gesture, the need for a mannequin capable of having the correct geometry, and the fact than if the authentication system is to open a door the need to deploy the mannequin is hard to do without arousing suspicion.

6.4 Description of our experiments

6.4.1 The Kinect^{TM} protocol for the experiment

Extracted features

We have chosen to recognise the gesture based only on the movement of the arms and hands, in order to be independent of the user position in front of the sensor: stepping or flexing would not be recognised i.e. the position of the feet and of the lower body is not relevant information. A particular advantage of using fewer points is that the system needs less physical space to operate in, as the user does not have to be far from the device.

From the 20 points that the Kinect$^{\text{TM}}$ provides at a time *t*, we extract the position of both hands, elbows and shoulders (six points in total).

Using just these 6 skeleton points we estimate that there are $\approx 10^{144}$ 1 second (30 frames) gestures: this assumes humans are only able to achieve a pose accuracy of $1cm^3$, giving $\approx 10^{16}$ potential gesture starting points and $\approx 2.5 \times 10^4$ possible positions in each subsequent frame). To achieve position-independence, we refocus all the positions of the limbs used onto the torso of the user as follows:

$$X_{refocus,t} = (X_{x,t} - T_{x,t}; X_{y,t} - T_{y,t}; X_{z,t} - T_{z,t}) \tag{6.1}$$

where:

- $X_{refocus,t}$ is the position of a limb *X* refocused on the torso at a moment *t*
- $X_{x,t}; X_{y,t}; X_{z,t}$ is the position, in x, y, z, of a limb X such as the Kinect$^{\text{TM}}$ gave us at a moment *t*
- $T_{x,t}; T_{y,t}; T_{z,t}$ is the position, in x, y, z, of the torso at a moment t

We divided the volunteers into two groups, a *reference group* composed of 10 people and an *attacker group* composed of 28 people.

In the reference group, our sample consisted of 4 women and 6 men: their height ranges from 1.51m to 1.83m and arm length ranges from 0.68m to 0.75m.

For the attack group, there were 13 women and 15 men, whose height varied from 1.53m to 1.92m, and their arm length was between 0.62m to 0.79m.

In the reference group, each person was asked to register a gesture of their own design, following some basic instructions. Gestures should be:

- relatively short, as our system records gestures of a maximum five second length;
- based on a movement of their hands and arms, as we refocus all points on the user's torso so we cannot recognise any body movement (e.g., step forward, flexing, etc.);
- easily reproducible (i.e. they have to pay attention to where their arms are at the beginning and end, etc.);
- and not too close to the body, as the Kinect$^{\text{TM}}$ has difficulty tracking the arms when they are too close to the body.

Once the volunteers had registered their gestures, they were asked to reproduce it 20 times. This allows for the calculation of the True Accept Rate (TAR) / False Acceptance Rate (FAR).

Reference gestures were also recorded by a separate camera in order to be shown to the attack group in the next phase of the experiment.

In the attack group, each volunteer was asked to try and guess the gestures of the reference group, without knowing what they were. To do that they were given 2 minutes to attempt all gestures which came to their minds.
We then showed the attacker a recording of a reference gesture and asked them to copy the gesture as accurately as possible 10 times: this was repeated for each reference gesture.
This allows the False Acceptance Rate / True Reject Rate to be calculated.

6.4.2 The Leap Motion Protocol for the Experiment

We set up a GR experiment, which produced 90,000 attacks and 10,000 attempts at authentication by genuine users.

Gestures were captured by a Leap Motion device which tracks and records hand movements in three dimensions. We recorded the (x, y, z) positions of the palm centre and all five fingertips and finger roots (i.e. eleven elements (E) for each frame).

A group of ten volunteers was asked to reproduce ten pre-determined gestures ten times.
It is important to note than this experiment was designed to simulate an attack. We thus disclosed to all participants the design of each gesture to be mimicked.

We devised a set of 10 model gestures ranging from a simple hand drop, to more complex shapes, e.g., drawing a symbol of infinity, and gave instructions about which hand to use and the positioning of fingers (see Figure 6.2).

Five gestures were done with an open palm, as follows:

01: Let Right Hand (RH) drop vertically.
02: Make a circle on horizontal plane with Left Hand (LH).
03: Make a square on vertical plane with RH.
04: Make a triangle on horizontal plane with LH.
05: Draw a symbol of infinity on vertical plane with RH.

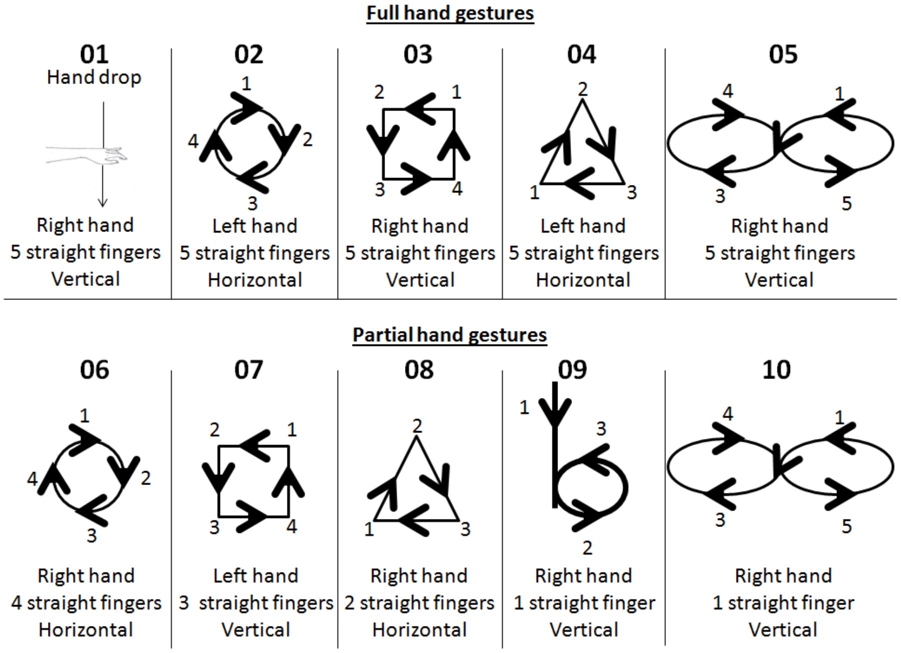


Figure 6.2: *Model Gestures (Designed: Benoit Ducray)*

Another five gestures required variation of finger position, as follows:

06: Make a circle on horizontal plane with the RH and the index, middle, ring and little finger straight.

07: Make a square on vertical plane with LH and the index, middle, ring finger straight.

08: Make a triangle on horizontal plane with LH and the index, middle finger straight.

09: Draw a ”b” on vertical plane with RH and the index finger straight.

10: Draw a symbol of infinity on vertical plane with RH and the index finger straight.

In order to accurately record when a gesture starts and stops, and to ensure that authentication attempts are synchronized with stored gesture templates, the software waits for an unmoving hand with five straight fingers before triggering or stopping the recording.
More synchronisation is done automatically during the analysis with the DTW algorithm.

6.5 Results

6.5.1 The Kinect^{TM} experiment results

Table 6.2: *False acceptance rate*
Success / number of attacks for each person in reference group.
**Weak gesture, excluded from calculations (source: Benoit Ducray).*

Person	False Acceptance rate	Person	False Acceptance rate
Person01*	34.7%	Person06	1%
Person02	0%	Person07	1%
Person03	5%	Person08	2.1%
Person04	3.54%	Person09	0%
Person05	2.8%	Person10	0%

By the end of the experiment we had obtained:

- 200 sample gestures of genuine users trying to authenticate;
- 56 minutes of people trying to guess the reference gestures (assuming each guessing attempt lasted an average of 3 second this equates to 11200 guessing attacks);
- 2800 attacks done by copying the reference gestures.
  Out of these 2800, only 142 attacks succeeded including the ones on the weak gesture (see below).
  However, we found out that 70% of the successful attacks occurred when attacking one particular gesture. We concluded that this gesture (a simple movement of both arms up and down) was a particularly weak gesture that could be excluded from the subsequent calculations as it may be considered equivalent to a weak PIN.
  In the end there were 2520 attacks done by copying the 9 remaining reference gestures.
  Out of these 2520, only 44 attacks succeeded, i.e. 1.7%.

Figure 6.3 presents the ROC curve, when an attacker knows the gesture.
From this curve, we can determine the best threshold which maximizes the TAR and minimises FAR. In our case this threshold would be 0.40, which gives a TAR of 93% and a FAR of 1.7%, shown in black in Figure 6.3.
Based on this threshold we can find the successful attack rate for each reference person (see Table 6.2).

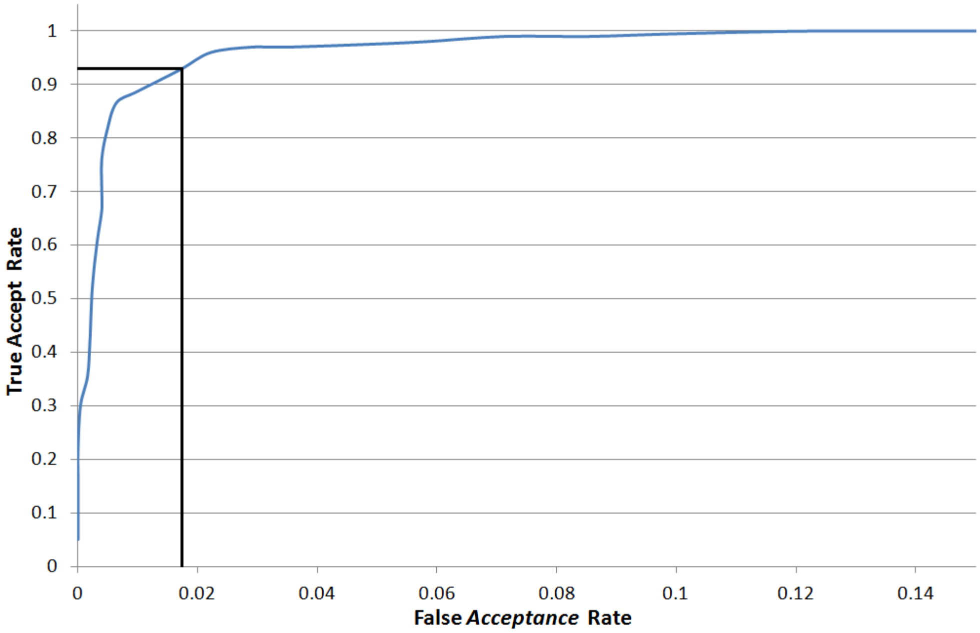


Figure 6.3: *ROC curve when the attacker knows the gesture (Kinect^{TM}).*
Black lines indicate a threshold of 0.4 giving TAR 93% / FAR 1.7% (source: Benoit Ducray)

From this ROC curve we can also determine the Equal Error Rate (EER).
In the case of an attacker knowing the reference gestures, the EER is 2.8%.

When the attacker does not know the gestures they do not have any choice other than to move randomly in front of the Kinect^{TM}.
Each volunteer did that during 2 minutes, and their movements were compared to all the recorded gestures with the threshold we have determined previously (0.40).

Assuming the average gesture duration produced by the volunteers in this case is 3 seconds, that means in 2 minutes they are able to make 40 attacks, so our volunteers produced 11200 attacks in 56 minutes.
None of these attacks had any success which make it impossible to draw a ROC curve or to determine the EER.

However, the conclusion that can be drawn here is that the likelihood of a successful attack is less than 1 in 11200 which gives better security than a 4-digit PIN which has 10,000 possible combinations.

6.5.2 The Leap Motion experiment results

We also tested GR using just the palm centre data points. This effectively removed all hand geometry information, resulting in similar data to that which could be obtained from an accelerometer sensor.

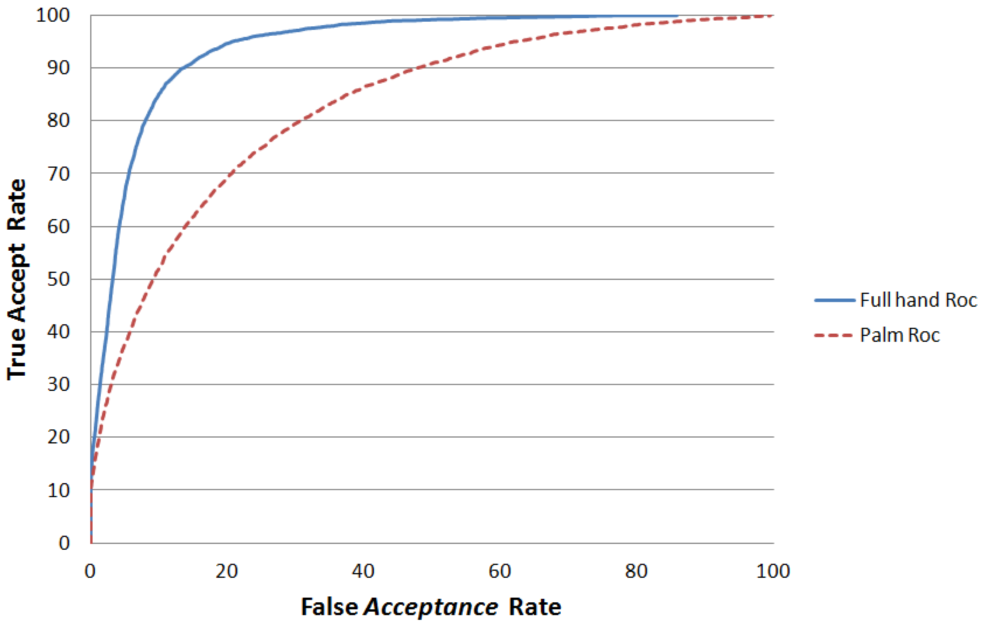


Figure 6.4: *ROC curves of the full hand gesture authentication and palm gesture authentication - Leap Motion (source: Benoit Ducray)*

The ROC curves obtained are shown in Figure 6.4. For full hand gestures, it can be seen that for a TAR of 10 we have FAR of 0; that is due to the comparison of a user's gesture with itself which should always give the minimal score, i.e. 0.

The asymptotic part takes a long time to reach the rate of 100, which means some of the samples are distant compared to the other samples from the same user's gesture. This may be due to several users having difficulties using the Leap Motion as it was the first time they had tried it.

Figure 6.4 shows that palm GR is a less effective biometric for secure authentication than the full hand gesture; the EER from the palm centre gesture authentication is 25.04% against 11.88% for the full hand. The latter rate shows that full hand gestures produce half the number of errors (i.e. False Rejection/False Acceptance) than a gesture authentication based on the palm alone.

From the analysis of the ROC curve, we determined the optimal threshold for both the full hand gesture and palm gesture, with the help of the EER. With these respective thresholds, we find the global TAR= 88.12% and the FAR 11.88% for full hand gesture and the palm gesture TAR of 74.96% and a FAR of 25.04%. Once we had determined these thresholds, the analysis was done gesture by gesture.

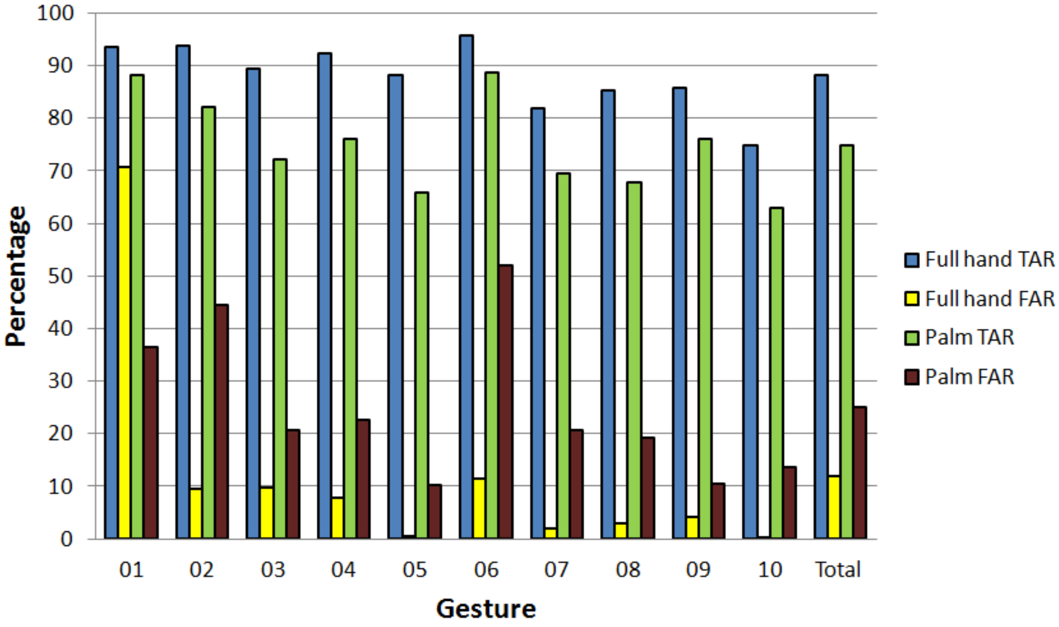


Figure 6.5: *Split of TAR and FAR per gesture for the full hand and palm GR (source: Benoit Ducray)*

Figure 6.5 shows the split of TAR and FAR per gesture for both hand and palm gestures. It can be seen that none of the TAR reach 100%.
This may be due to the fact that the users are not familiar with using the Leap Motion. 88 Also most of the gestures without straight fingers had lower TAR than the other gestures, which suggests that the users were not at ease performing this kind of gesture.
In particular, gesture 10 had a very low TAR of 74.8%, possibly because it was the most difficult and potentially less comfortable for the users to execute consistently.

The most basic gesture (hand drop 01) corresponds to simple analysis of only the user hand geometry and gave the worst result: the FAR is 70%. As we examine more complex gestures the FAR is never higher than 10%.

This point is confirmed by gesture 05, which is the most complex of the full hand gestures, and which has 0.47% of TAR. Analysing Figure 6.5 shows that for every gesture, the TAR is always better for the full hand GR (except for the outlier gesture 01).
We could therefore deduce that for authentication based on GR, it is better to have the most information possible from the sensor as more points captured from the hand would give better results.

As a result, we can say that an attacker may be able to copy a gesture exactly after practising it several times. However the biometric hand geometry of the user will still afford some protection in this case.

6.6 Contextualization of our results

6.6.1 Comparison with other research works

Table 6.3: *Comparison the results of our experiments with those provided in several gesture authentication papers.*
FA-BF: False Acceptance brute force, FA-AK: False Acceptance attacker knows
(source: Benoit Ducray)

Papers	Sensor	Algorithm	TAR	FA-BF	FA-AK
Our Leap Motion experiment	Leap Motion	DTW	88.12%	–	11.88%
Chahar et al. [26]	Leap Motion	Mix of Naive Bayes, Neural Network, Random Decision Forest	81%	1%	–
Aslan et al. [14]	Leap Motion	DTW	88.29%	–	11.71%
Aumi et al. [15]	Short range depth sensor	DTW	96.6%	3.4%	5.3%
Our Kinect™ experiment	Kinect™	DTW	93%	0%	1.7%
Wu et al. [117]	Kinect™	DTW	98.11%	–	1.89%
Tian et al. [108]	Kinect™	DTW	99%	1%	3%

In Table 6.3 we gathered the results of our experiments with those of several papers published on the GR authentication subject, some of which using protocols different from the ones we followed.

Wu et al. [117] proposed to use all 20 skeleton tracking points provided by the Kinect^{TM} and it gave them a TAR of 98.11% for 1.89% of FAR.

Tian et al. also used the Kinect^{TM} with the DTW algorithm for analysis and recognition of gestures designing a 3D signatures [108]. In this work they found 99% of TAR for 1% FAR and 3% against attacks.

Aumi et al.'s paper [15] explored hand gesture authentication, accuracy and attack resistance against shoulder surfing. In this experiment, reference hand gestures were recorded using a depth camera, filmed, and shown to a group of attackers. Participants were then asked to copy given gestures [15]. In this case, the FAR was 2.3%.

We can thus see that the papers compared in table 6.3 show that TAR can vary between 81% to 99% depending on the capture system used.
The table also shows results of brute force attacks against these systems (denoted FA-BF), where attackers attempted to guess a gesture.
It can be seen that this type of attack is very unlikely to succeed.

Additionally if the attacker knows the gesture (denoted FA-AK) the FAR results vary from 1.7% to 12%. Our results confirm the other experiments done on this domain by researchers.

More generally, studies based on Kinect^{TM} get better results than the one based Leap Motion or on short range depth sensors, which may imply that the more parts of the body that are used for the authentication, the better it is for correctly accepting the genuine user and the rejection of potential attackers.

6.6.2 Comparison to fingerprints

In order to extent the analyse on GR, we will compare our work with Fingerprint recognition, as it has been well proven and a widely used biometric [53].

To compare these two methods, we will study two security abilities of the system: the ability to correctly recognise a genuine user, and the ability to reject an individual trying to mimic the biometric.

- For the recognition of a genuine user, either the system is performing an identification or an authentication. In this case, GR cannot compete with fingerprint recognition. Fingerprint recognition has an error rate of 0.2% [60] while for the best GR system it is around 1% or 2%. We see that fingerprint recognition is a better biometric system than other presented to correctly recognise a genuine user.

Table 6.4: *Fingerprint liveness detection algorithm performance (data extracted from [47])*

Fingerprint Liveness detector FAR		
Hardware used	Algorithm	FAR
Sagem sensor	Local Binary Patterns	4.34%
Digital sensor	Valleys wavelet	12.40%
Digital sensor	Wavelet energy	15.10%
Sagem sensor	Ridges wavelet	18.15%
Sagem sensor	Power spectrum	21.81%
Italdata sensor	Pores detection	22.00%

- In the case of a malicious individual mimicking the biometric in order to fraud the system, we do not take into account the ability of GR of being changeable as it is a dynamic biometric. We compare the FAR found for the GR to the FAR of liveness detection for fingerprint.
  Several works have been gathering the results of experiments related to liveness detection such as [47, 102].
  Table 6.4 presents some of the best liveness performance algorithms found by [47].

  For GR, the FAR found, either by our experiments or others works, is around 2% or 3% for systems based on the Kinect^{TM} and around 12% for systems based on the Leap Motion. These rates are better than most of the results presented in Table 6.3.

  But it is also important to take in consideration what [102] said about the liveness detection, i.e. “The results suggest that the performance of the methods strongly depends on the knowledge of the fake fabrication techniques and materials during the development of the method.”
  In other words, this means that liveness detection requires the authentication authorities to keep up with the evolution of the forging techniques and materials (e.g., in this case the fabrication of a synthetic finger).
  FAR would largely vary according to the kind of attack and the knowledge of the techniques deployed.

  Despite the apparent results shown here above, GR is finally a better authentication system to use as the method used to forge biometrics evolve continuously and the one that will be used for the next serious attack is potentially unknown yet.

6.7 Summary

In this chapter, we have presented four kinds of attack from which a GR system may suffer one of the following attack: “Brute Force attack”, “Dictionary attack”, “Storage Leakage” and “Shoulder Surfing attack”.
We have focused on two of these attacks, viz. “Brute Force attack” and “Shoulder Surfing attack”, by setting up two experiments which used different devices to capture the gesture:

1. the Kinect$^{\text{TM}}$, which tracks the full body movement
2. the Leap Motion, which tracks movement of the hands.

The results of the Kinect$^{\text{TM}}$ experiment, in the case when the attacker does not know the users secret gestures, show that none of the attacks succeeded during 56 minutes of attack (over 10,000 attacks), which is at least as good as an unknown 4-digit PIN. When the attacker knows the users gestures, however the system has 1.7% FAR for 93% of TAR and an EER of 2.8%, which remain a rather low attack success level, even if significantly above the scores obtained with fingerprints and retina.

For the Leap Motion experiment, we chose eleven characteristic points, i.e. the palm centre and all five fingertips and finger roots. We used the DTW algorithm to compare the gestures. This experiment was more focused on a mimic attack as all participants had to execute predetermined gestures.

An attacker mimicking a known gesture had 11.88% likelihood of a successful attack, whilst a genuine user had a 88.12% chance to be correctly authenticated.

All these results align with others’ studies done on this domain, using the same kind of sensor.

A comparison was made with fingerprint recognition (FAR scores) where it has been established that for recognizing a genuine user, either for identification or authentication, fingerprint is much better than GR.
However, against an attack GR is better for rejecting an attacker unless we know the method used to forge the biometrics. Should the forging technique be known, the liveness detection algorithms for fingerprints obtain a better score but in the real world such knowledge is greatly unlikely.

Chapter 7

Gesture Recognition Implemented on a Personal Limited Device

Contents

This chapter is mainly based on our paper [41], which was nominated for the best paper award at the International Conference on Information and Communication Systems (ICICS 2017).

The main challenge of a biometric authentication system is to protect the integrity of the reference template. Should an attacker obtain or tamper with the template, the genuine user would have to stop using that biometric for any application.

This chapter investigates the feasibility of implementing a Gesture Recognition (GR) system on a personal limited device such as a smart card. To do this, we set out an experiment using sample gestures based on practical results of gesture authentication trials and an optimised version of Dynamic Time Warping (DTW) algorithm to analyse the data captured. We implemented them on both a contact Smart Card (SC) and a smartphone, a much more powerful device, using Host Card Emulation (HCE). The latter being 60 times quicker than the former (one second vs one minute).

7.1 Problem statement

One of biometric systems' principal challenges is to store the biometric reference template in a secure location. If a malicious individual is able to obtain the template, that would mean the genuine user would not be able to reuse this biometric for any application.
A prime location for the related reference template would be on a security evaluated Smart Card (SC), as it is tamper resistant, easy to carry and, if used with Gesture Recognition (GR), the system can provide a three-factor authentication method.

If the matching processes could also be carried out on-card (Match-on-Card), then this provides additional protection, as the template does not need to leave the card during an authentication. It also protects against attacks on the implementation due to the tamper-resistance of the smart card chip.

GR is a rather demanding system in terms of computation power and memory storage.
This chapter sets out to investigate whether it is feasible to implement a GR system on a personal device of limited capability, such as an SC. To do this, an experiment was performed using sample gestures based on practical results of gesture authentication trials which used depth cameras as sensors and the Dynamic Time Warping algorithm (DTW) [63] to analyse the captured data. We varied the data length, number of frames and tracking points of the sample gestures, and implemented them on both a contact SC and the much more powerful Samsung Galaxy S4 mobile phone. The latter used Host Card Emulation (HCE) [89] to emulate the SC, and the DTW algorithm was optimised to minimise memory usage on both platforms.

The experiment showed that the implementation on a SC was slow (in a excess of minute), and the HCE version was much faster (around 1s or 2s), although the overall processing time depended on the gesture data length. It should be noted that the test applications were implemented at the platform level, rather than in low-level native code which would have been much faster.

This chapter is structured as follows: Section 7.2 presents some background information about Host Card Emulation. In Section 7.3, we present the details of the experiment, gestures, hardware and the optimisation of the gesture comparator used. The Section 7.4 present the results. In Section 7.5 we discuss the feasibility of using GR on a personal device of limited capabilities.

7.2 Host Card Emulation

Host Card Emulation (HCE) is a technology which emulates an SC on mobile equipment using only software [89], and can be used with Near Field Communication (NFC) to emulate a contactless SC.

Before HCE, all messages from a card terminal were routed to a hardware Secure Element (SE) in the mobile handset. HCE communicates directly with the mobile operating system, which decides if messages should be handled by a physical SE or a software application [111].

7.3 Experiment

In the previous chapter, we used six of the twenty available Kinect$^{\mathrm{TM}}$ skeleton-tracking points (we wanted the use to be able to seat and get authenticate so we chose to focus on both hand elbow and should which should be visible by the device even seated) in a gesture authentication system with promising results: giving an Equal Error Rate (EER) of 2.8%.

We devised a proof-of concept authentication experiment using a Leap Motion device, which tracks and records hand movements in three dimensions. For more technical information concerning the Leap Motion device see [3].

Preliminary results from a small sample of volunteers indicated that it is feasible to use this device in gesture authentication systems although the EER is 11.88%.

For the performance evaluation in this chapter, we used the Leap Motion to record a gesture of 90 frames (as we want a good sequence of the gesture) with 11 tracking points (the five finger tips and roots and the palm centre which are the most characteristic point of the hand) from which we truncated the floating numbers and encoded them all into two bytes.

We chose to emulate gestures from these two capture devices, setting the number of frames and tracking points in our sample gestures accordingly to reflect the different characteristics of the sensors. The DTW algorithm was used to analyse the gestures.

We are not assessing here the performance of any cryptographic protocols because they would be the same for both SC and HCE.

Table 7.1: *Information on the gestures and APDU sent (source: Benoit Ducray).*

	Gesture size in bytes	Number of frames	Number of tracking points	Number of frames sent per APDU	Size of the APDU data	Number of APDU sent
Gesture 1	3240	90	6	1	36	90
				6	216	15
Gesture 2	1764	49	6	1	36	49
				6	216	9
Gesture 3	5940	90	11	1	66	90
				3	198	30
Gesture 4	3234	49	11	1	66	49
				3	198	17

7.3.1 Gesture data and hardware

The gestures

We created four different sample gestures with varying memory requirements and processing time, described as follows:

- **Gesture 1**: This gesture represents the capture of six points in three dimensions and is composed of 90 frames. The total amount of data of this gesture is 3240 bytes. This gesture represents a three second gesture obtained with a device capturing at 30 frames per second. The number of tracking points represents either the five fingertips and the palm centre (if performing hand GR), or both hands, elbows and shoulders for upper body GR.

- **Gesture 2**: This gesture captures six points in three dimensions and is composed of 49 frames. The total amount of data of this gesture is 1764 bytes.
  This gesture may represent a 1.63s gesture obtained with a device capturing at 30 frames per second (it would be different with a device capturing more or less frames per second). The tracking points are the same as in Gesture 1.

- **Gesture 3**: This gesture captures 11 points in three dimensions and is composed of 90 frames. The total amount of data of this gesture is 5940 bytes. Again, this gesture may represent a three second gesture obtained with a device capturing at 30 frames per second. The tracking points can represent the five fingertips, five finger roots and the palm centre for hand GR, or both feet, knees, hands, elbows, shoulders and the head for body GR.

- **Gesture 4**: This gesture captures 11 points in three dimensions and is composed of 49 frames. The total amount of data of this gesture is 3234 bytes. This gesture may represent a 1.63s gesture obtained with a device capturing at 30 frames per second. The tracking points are the same as in Gesture 3.

The hardware

The devices used for the experiment were: an ACR1281U reader which can be used with both SC and NFC devices as contactless reader, attached to a PC running Windows 7 with 2 GB of RAM and a processor of 1.86 GHz. As an SC, we used a Java Card 2.2.2 with 16 bits processor, and a HCE equivalent application running on a Samsung Galaxy S4 with Android 5.0.1, 2 GB RAM, Quad-core (4x1.6 GHz Cortex-A15 and 4x1.2 GHz Cortex-A7).

The experiment protocol

Firstly, we needed to decide how to send the gesture information from the terminal to the card. A normal Application Protocol Data Unit (APDU), which is how we communicate with a card, can send up to 256 bytes. We tested two methods for sending the gesture information:

- Sending all the information frame by frame: in this way the APDU data size is 36 bytes for Gesture 1 and 2 and 66 bytes for Gesture 3 and 4
- The other method was to send the maximum number of frames that an APDU can handle. For Gesture 1 and 2, it is six frames which gives an APDU data size of 216 bytes and for Gesture 3 and 4, it is three frames, so the APDU data size is 198 bytes

All the information about the gestures and the APDUs sent are summarized in Table 7.1.

We measured the communication time for the APDUs described above for both SC and HCE, in order to know how this decision may affect the performance evaluation. We carried out 100 time measurements, to assess if the measured response time was stable.

We then performed the GR application with DTW. First, we captured 100 time measurements for each of the four gestures, when running the application by sending the gesture frame by frame to the SC. We repeated the experiment packing the maximum number of frames into the APDU. We then repeated these two steps using HCE.

7.3.2 Dynamic time warping: memory optimisation

The main drawback of the DTW algorithm is that, for a gesture A of M frames and a gesture B of N frames, it needs to fill an M x N matrix where the cell (i,j) represents the score between frame i of gesture A and frame j of gesture B plus the cumulative score.

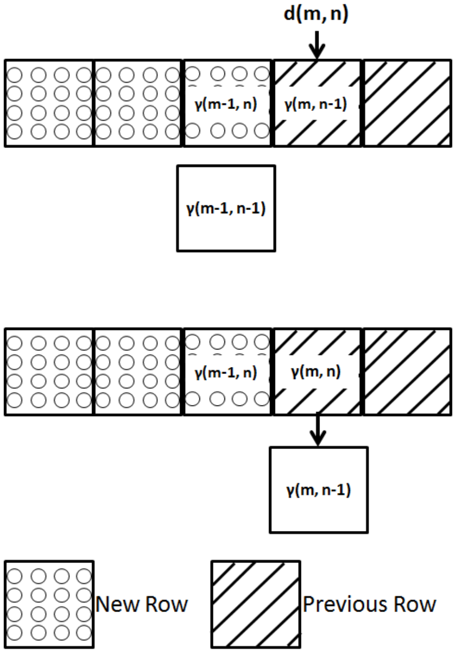


Figure 7.1: *Application of DTW with only one row in memory (source: Benoit Ducray)*

Some works try to optimise the DTW either in calculation, memory or both. In [90], they reduced the amount of calculation and memory needed by focussing on a part of the DTW matrix, which may contain the warping path. But they forced the warping path of any comparison to be in this calculated section which may imply more false positive results. The same comment can be made if we do not calculate the full matrix as the warping path will be altered.

Equation 4.3 shows that we only need three elements, $\gamma(m-1;n-1), \gamma(m-1;n), \gamma(m;n-1)$. Thus we only need to have in memory two rows, either the row m and row m-1, or the row n and row n-1. Let us say that we have in memory the row m and row m-1: this method then reduces the memory cost from M x N to 2M, although the number of calculations remain unchanged.

We found that it is possible to implement the DTW algorithm by storing only one row of size M plus a temporary variable of the size of one element of M. This method overwrote the row at each iteration and saved the temporary variable in the last overwritten cell.

So if we are looking for $\gamma(m;n)$ which will be saved in the cell c(a), we can find $\gamma(m;\ n-1)$ in the cell c(a-1); $\gamma(m-1;n)$ is the current value of c(a), and $\gamma\ (m-1;n-1)$ is saved in the temporary variable.

Once the cost $\gamma(m;n)$ is calculated, we have to save the value in c(a) in the temporary variable then overwrite c(a) with the value of $\gamma(m;n)$. Figure 7.1 is illustrating this method.

7.4 Results

For the communication, the SC is always more than three times quicker than the HCE (the average time for each kind of APDU can be seen in the Table 7.2 in the communication time per APDU column) and is more stable as its average standard deviation is 0.29ms against 5.65ms for HCE.

Table 7.2: *Times measured in millisecond (source: Benoit Ducray)*

	Size of the APDU	Number of APDU sent	Communication time per APDU		Average time for the full application		Estimate processing time	
			SC	HCE	SC	HCE	SC	HCE
G1	36	90	6.71	23.65	92982.15	2228.48	92378.09	99.48
	216	15	24.72	77.54	76029.02	1196.35	75658.16	33.12
G2	36	49	6.71	23.65	50622.60	1214.52	50293.73	55.40
	216	9	24.72	77.54	43760.30	667.18	43555.80	23.14
G3	66	90	9.79	31.81	108674	3084.76	107792	221.60
	198	30	23.76	71.80	94083.19	2253.42	93370.15	99.30
G4	66	49	9.79	31.81	32448.10	1684.36	31968.07	125.52
	198	17	23.76	71.80	28092.53	1237.19	27702.44	64.68

We then measured the time for the full application (communication time plus processing time) as described earlier in 7.3.1. The average time for each gesture can be seen in Table 7.2 under the column Average time for the full application.

Knowing the communication time, the number of APDUs sent and the full time for the application, we can calculate the time needed by the devices to process the GR. This time can be seen in Table 7.2 under the column Estimate processing time.

An extended version of the Table 7.2 can be found in Annex B.

We observed that the SC needed a lot of time to process a gesture; more than 27s for the quickest. Even if we used SC technology with a quicker communication interface, the SC processing would still be a bottleneck. On the other hand, the HCE had a slow communication time, but its processing time was much quicker than the SC, rarely exceeding 100ms duration.

7.5 Discussion

The experiment has shown that it is possible to implement authentication based on GR on an SC at platform level, however the performance (one minute duration for a three second gesture) was far too slow to be practical. A solution to this problem could be to develop the application on a lower level, either in hardware or in native code.

Based on the work of [52] who implemented signature recognition on an SC, on both application level and native level, using an algorithm of similar complexity to the one we used. We estimate that the process time would be three times faster.

An HCE application would be more feasible for a real application as it takes around one second. However, an HCE application does not provide the attack resistance offered by an SC. The HCE application could be protected, at least from phone malware, by running within a Trusted Execution Environment (TEE).

A TEE offers a more restricted and protective environment for running sensitive code, compared to normal phone applications [45], although it does not offer the tamper resistance of an SC.

Using a device with faster communication speed, or devices supporting extended APDUs (an extended APDU is able to support up to 2^{16} data bytes [1]), will reduce the time needed for the full GR application.
The application installed on a personal, limited device will still remain slower than an equivalent application installed on a secure server, but it will be more versatile, supporting both on-line and off-line transactions.

An example application that could use this kind of authentication is controlling access to a building. If a sensor and a reader are installed at a restricted area entry point, possession of the SC (or the phone) plus knowledge of the correct gesture performed in the correct manner would be needed to enter. This three-factor authentication then reduces the likelihood that the system could be compromised.

7.6 Summary

In this chapter, we pondered on a secure location to store a GR template. We investigated whether an SC or an HCE implementation would be feasible for a GR application, when using the DTW algorithm to compare gestures. Thus, we measured communication and processing time for both SC and HCE.

Although it is possible to run a GR application on an SC at platform level, it is not feasible for a real application as our implementation took around a minute.
There is margin for improving performance by implementing the application at a lower level, either in hardware or with native code.

Our HCE application was far more practical for a real world application, although it did not provide the attack resistance that an SC offers.
The use of a TEE may enhance security and resist logical and malware attacks, although a TEE does not offer the tamper-resistance of an SC.

Chapter 8

Application and threat model

Contents

In the previous chapters, we shown that authentication application based on authentication using Gesture Recognition (GR) is a good authentication system. We also shown that a Smart Card (SC) secure place a template and run the authentication application. But all of these is useless if any malicious individual can spy, inject or modify the communication between the SC and the terminal.

In this chapter we propose a security protocol in order to protect the transmission of a gesture from a terminal to a SC for an authentication application based on authentication using GR. We assess the security of this protocol through Scyther, which is a formal analysis tool for security protocols.

We described different types of attackers: Script Kiddies, Expert individuals and Organised crime. We determine the abilities of each of them to attack the system. We deduce from this analysis that most of the profiles do not have more possibilities of bypassing the authentication other than just trying to mimic the gesture.

8.1 Problem statement

Chapter 6 has shown that an authentication based on Gesture Recognition (GR) is able to efficiently recognize the genuine user and reject any attackers or malicious individuals.

We have also seen in chapter 7 than such authentication can be versatile by supporting both on-line and off-line transactions, as the gesture template can be securely stored on an SC and the authentication can be run on the same device.

A good authentication system and a secure place for the template is useless if any malicious individual can spy, inject or modify the communication between the SC and the terminal.
We thus need a security protocol which makes it nearly impossible to affect the confidentiality, the integrity and availability of the authentication data.

In this chapter, we propose a security protocol between the SC and the terminal, in order to assess whether the whole system (gesture authentication and the security protocol) correctly authenticates the genuine user while keeping the gesture secret.

We identify four kind of attackers' profile: Script Kiddies, Expert individuals and Organised crime. We determine the abilities of each of them to attack the system based of the threat vectors presented in Chapter 3.3.
The present research shows that most of the profiles do not have more possibilities of bypassing the authentication other than just trying to mimic the gesture.

This chapter is structured as follows: Section 8.2 presents the entities involved for the security protocol and assumptions we make around the security protocol and the minimum security goals of the protocol. In Section 8.3, we present the details of the security protocol we propose and assess the protocol's goals. In Section 8.4 we present different of attacker profiles. Section 8.5 describes different kinds of attacker profiles and attacks which cover the threat vectors presented in 3.3. In Section 8.6 we discuss the attack possibilities of each profile.

8.2 Background on the security protocol

8.2.1 The entities

For a security protocol, there are several entities: one or several authorities, a terminal, and an SC.

- *The authorities*: the role of the authority is to certify one or both devices (i.e. the SC and terminal). This certification is used to assert the authenticity of the devices.

  For our application, there is only one authority which has certified both the SC and terminal.

- *The terminal*: The terminal consists of a smart card reader, a depth sensor to capture gestures, and a computer. Its role is to be an interface between the sensors and the SC. Moreover, in terms of the whole authentication application, the terminal is the interface between the user and the entity with whom they want to do a transaction.

  For our application, the terminal needs to have an identification (ID) number through which it can be identified (noted Id_T), a couple of RSA keys [93] (SK_T for the private Key, PK_T for the Public Key), and a digital certificate which has been signed with the secret key of the authority ($S_{SK_A}\{PK_T\}$). The terminal also has the public key of the authority (PK_A) in order to verify the authority signature.

- *The Smart Card*: The role of the SC is to store the gesture template and to run the gesture comparison algorithm in order to determine if the attempted gesture comes from a genuine user or from an impostor.
  For our application, the SC needs to have the gesture template, an ID number through which one it can be identified (noted Id_{Sc}), a couple of RSA keys [93], and a digital certificate which has been signed with the Secret Key of the authority ($S_{SK_A}\{PK_{SC}\}$). The SC also has the Public Key of the authority (PK_A) in order to verify the authority signature.

8.2.2 Assumptions

Before going any further, we need to state a series of assumptions:

- We will consider the SC to be trustworthy, due to the properties of the SC's integrated circuit with some form of tamper resistance.

- We will consider the terminal to be trusted. We are making this assumption as we consider that the terminal should be in a controlled environment.

- We will consider all cryptographic keys to have been checked for validity before use and that adversaries cannot break cryptographic algorithms.

8.2.3 Minimum security goals of the protocol

- Mutual Entity Authentication: Both of the entities, the SC and the terminal, authenticate to each other. This goal is to avoid masquerading by a malicious entity.
- Public Key Exchange: In order to facilitate the key generation and the entity authentication process, an exchange of certified public keys has to be done between the entities.
- Key Freshness: The generation of a key has to be fresh to the protocol session, in order to avoid replay attacks.

8.3 Proposed security protocol

The security protocol between the terminal and the SC is shown in Figure 8.1 and the notation used is in Table 8.1.

1. The terminal sends the SC its public key (PK_T) in the clear, its digital certificate signed with the secret key of the authority ($SK_A\{PK_T\}$), its identity number (Id_T), and a first nonce (n_{T_1}).

2. The SC checks, using PK_A, that the terminal digital certificate it has just received has been signed by the authority.

   If the signature is correct, the SC generates two 192 byte keys. One is used as an AES session key (*K*) to encrypt subsequent messages. The other is an AES key (K_{MAC}) which will be used to generate the Message Authentication Code (MAC) by a CBC-MAC method.

   Then the SC sends back a message to the terminal which is encrypted with PK_T. This message is composed of the SC public key (PK_{SC}) and its digital certificate also signed with the secret key of the authority ($SK_A\{PK_{SC}\}$), the Starting Variable (SV) for the Cipher Block Chaining (CBC) mode encryption, the Starting Variable for the CBC-MAC (SV_{MAC}), the session key (*K*), the MAC key (K_{MAC}), the Id_T, the SC identity number (Id_{SC}), n_{T_1} and an SC generated nonce (n_{SC_1}).

3. The terminal first decrypts the message using the SK_T and checks that n_{T_1} and the terminal identity number it has just received is the same one that it previously sent. If the nonce and the ID are correct, the terminal will verify the digital certificate signature using PK_A.

Table 8.1: *Notation table (source: Benoit Ducray)*

Notation	definition
Entities	
A	Authority
T	Terminal
SC	Smart Card
Keys	
SK_Z	Private key of the entity Z
PK_Z	Public key of the entity Z
K	A 192 bits AES key session
K_{MAC}	AES key of 192 bits which will be use to generate a MAC
SV	Starting Variable for the CBC mode of encryption
SV_{MAC}	Starting Variable for CBC-MAC generation
Fields	
$S_{SK_Z}\{X\}$	The element X has been signed with the private key of the entity Z
$e_Y(X)$	The message X has been encrypted by an asymmetric encryption method using the public key Y
$E_Y\{X\}$	The message X has been encrypted by an symmetric encryption method (AES using the CBC mode) using the key Y
$MAC_Y < X >$	A MAC has been computed on the data X using a CBC-MAC method with the key Y
Other protocol elements	
ID_Z	Identification number of entity Z
n_{Z_i}	The i^{th} nonce generated by the entity Z
$Gesture_i$	The i^{th}fraction of a Gesture to be used in authentication

If the signature is correct, the terminal will start to send the gesture: this generates several messages. These messages are composed of an encrypted part which has been encrypted with the session key *K*, (using the AES as cryptographic function and the Cipher Block Chaining (CBC) mode) and a MAC part.

The encrypted part is composed of the first fraction of the $Gesture_1$, the Id_T and the Id_{SC}, n_{SC_1} and a newly generated nonce (n_{T_2}) which is encrypted with PK_{SC} in order to verify that the entity claiming to SC has the SK_{SC}. Based on this message, a MAC is generated with the key K_{MAC}.

4. Upon receipt of this message, the SC verifies that the MAC corresponds to the message. Then, after decrypting the message using the key *K*, the SC checks that n_{SC_1} is the one it sent previously, the Id_T and the Id_{SC}.

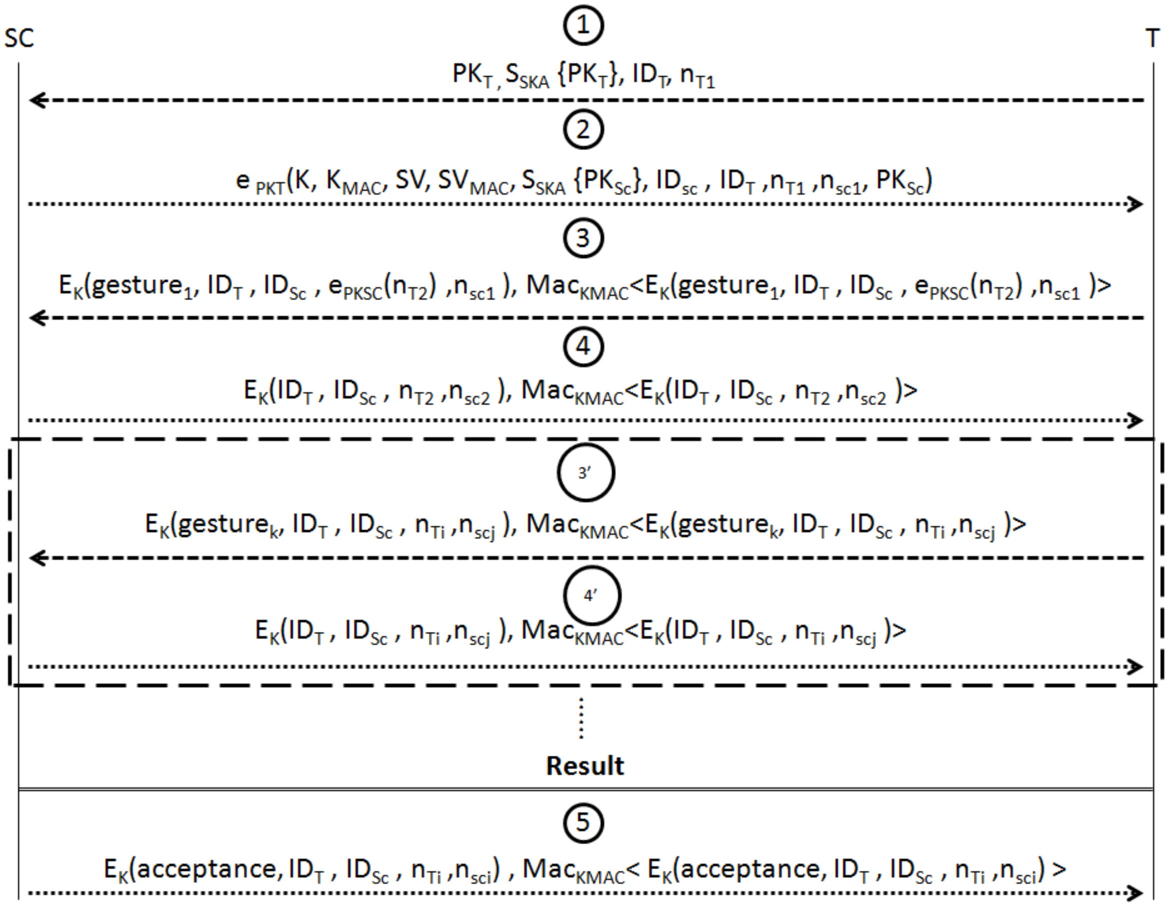

SC
T
1
$PK_{T,} S_{SKA} \{PK_T\}, ID_T, n_{T1}$
2
$e_{PKT}(K, K_{MAC}, SV, SV_{MAC}, S_{SKA} \{PK_{Sc}\}, ID_{sc}, ID_T, n_{T1}, n_{sc1}, PK_{Sc})$
3
$E_K(gesture_1, ID_T, ID_{Sc}, e_{PKSC}(n_{T2}), n_{sc1}), Mac_{KMAC}<E_K(gesture_1, ID_T, ID_{Sc}, e_{PKSC}(n_{T2}), n_{sc1})>$
4
$E_K(ID_T, ID_{Sc}, n_{T2}, n_{sc2}), Mac_{KMAC}<E_K(ID_T, ID_{Sc}, n_{T2}, n_{sc2})>$
3'
$E_K(gesture_k, ID_T, ID_{Sc}, n_{Ti}, n_{scj}), Mac_{KMAC}<E_K(gesture_k, ID_T, ID_{Sc}, n_{Ti}, n_{scj})>$
4'
$E_K(ID_T, ID_{Sc}, n_{Ti}, n_{scj}), Mac_{KMAC}<E_K(ID_T, ID_{Sc}, n_{Ti}, n_{scj})>$
Result
5
$E_K(acceptance, ID_T, ID_{Sc}, n_{Ti}, n_{sci}), Mac_{KMAC}< E_K(acceptance, ID_T, ID_{Sc}, n_{Ti}, n_{sci})>$

Figure 8.1: *Security protocol of communication between SC and terminal for authentication based on Gesture Recognition (source: Benoit Ducray)*

If everything is in order, the SC starts the gesture comparison and acknowledges successful receipt of the message by sending back an encrypted message composed of the Id_T and the Id_{SC}, n_{T_2} and newly generated nonce (n_{SC_2}). Based on this message a MAC is generated with the key K_{MAC} and sent with the message.

5. These two last steps are repeated (at the exception of the nonce encryption which no more encrypted with PK_{SC}) until the exchange gets to the last APDU containing the gesture data.

   When the SC receives the instruction of that last APDU, the SC checks if the MAC, the nonce, the terminal identity and the SC identity are valid.

   Then the SC finishes the gesture comparison and sends an encrypted message to the terminal containing the acceptance or the rejection, n_{T_i} and newly generated nonce (n_{SC_i}), Id_{SC}, Id_T and a MAC based on this message generated with K_{MAC}.

If any of the protocol validation checks fail, an error message is sent to all participants, the transaction is terminated and logged as unsuccessful.

8.3.1 The enrolment

Enrolment is when the system captures the reference biometric and extracts the features.
It is important that this phase is done in a secure environment where the identity of the user is checked in order to avoid false enrolment.

So, every time the user wants to enrol a new gesture, they have to go to the authority office which is considered secure and trusted.
A staff member will certify the identity of the user by asking to inspect the identity card and some basic questions about the user. Then the user will be allowed to record a new gesture and store it on the card.

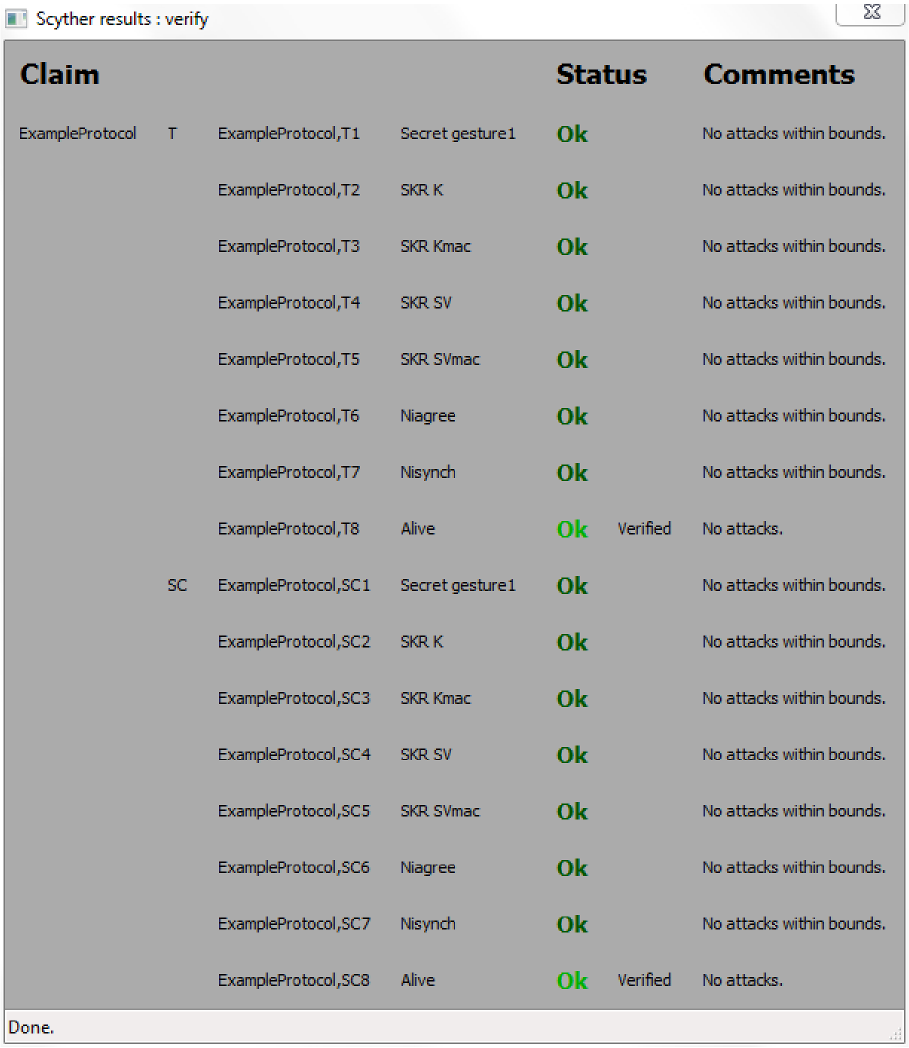
8.3.2 Protocol goals assessment

- Mutual Entity Authentication: To assess this goal, we will study how each entity authenticates the other individually.
  - SC authenticating the terminal:
    In message 1 the terminal sends its certificate and the PK_T from the terminal. The SC does not know where this message came from and so cannot trust it, but it can verify than the certificate is valid and has been issued by A using $S_{SK_A}\{PK_T\}$.

    The SC replies by challenging the terminal to decrypt a message encrypted with PK_T. This message is composed of two session keys *K* and K_{MAC}.
    If the terminal is the one it claims to be, it will be able to decrypt the message with SK_T and get access to *K* and K_{MAC}. Then the terminal replies with a message MAC section generated with K_{MAC}.
    By verifying the MAC, the SC has authenticated the terminal.
  - Terminal authenticating the SC:
    In message 2 the SC sends the terminal its certificate, Id_{SC}, and PK_{SC}. All these elements are encrypted with PK_T which has been certified by A.
    In order to prevent any leak of the SC certificate, at the same time that the terminal sends the first piece of gesture, it also sends its new nonce encrypted with PK_{SC} and expects the SC to send it back decrypted, in this way this the entity claiming to be the SC is proving that it has the SK_{SC}.

- Public Key Exchange: During the protocol each entity sends its certified public key to the other to help the authentication.
- Key Freshness: Both of the key sessions, K and K_{MAC} are generated by a random bit generator, which follows the recommendation of [17] and provides fresh entropy bits of a length of 192 bytes.

8.3.3 Formal analysis



Scyther results : verify

Claim				Status		Comments
ExampleProtocol	T	ExampleProtocol,T1	Secret gesture1	Ok		No attacks within bounds.
		ExampleProtocol,T2	SKR K	Ok		No attacks within bounds.
		ExampleProtocol,T3	SKR Kmac	Ok		No attacks within bounds.
		ExampleProtocol,T4	SKR SV	Ok		No attacks within bounds.
		ExampleProtocol,T5	SKR SVmac	Ok		No attacks within bounds.
		ExampleProtocol,T6	Niagree	Ok		No attacks within bounds.
		ExampleProtocol,T7	Nisynch	Ok		No attacks within bounds.
		ExampleProtocol,T8	Alive	Ok	Verified	No attacks.
	SC	ExampleProtocol,SC1	Secret gesture1	Ok		No attacks within bounds.
		ExampleProtocol,SC2	SKR K	Ok		No attacks within bounds.
		ExampleProtocol,SC3	SKR Kmac	Ok		No attacks within bounds.
		ExampleProtocol,SC4	SKR SV	Ok		No attacks within bounds.
		ExampleProtocol,SC5	SKR SVmac	Ok		No attacks within bounds.
		ExampleProtocol,SC6	Niagree	Ok		No attacks within bounds.
		ExampleProtocol,SC7	Nisynch	Ok		No attacks within bounds.
		ExampleProtocol,SC8	Alive	Ok	Verified	No attacks.

Done.

Figure 8.2: Formal analysis result (source: Benoit Ducray).

In order to do a formal analysis, we choose to use software called Scyther.
Scyther is a formal analysis tool for security protocols under perfect cryptography assumption.

This tool is usually used to find problems that arise from the way the protocol is constructed.

We submitted a simplified version of the protocol to this analysis. This simplified version goes directly from message 3 to message 5. This is done because messages 4, 3' and 4' are there to ensure than the whole gesture has been sent.
The code submitted to Scyther can be found in Annex A.

Scyther found that no attack is possible. The Scyther result is shown in Figure 8.2.
For both the SC and the terminal, we have tested that:

- The gesture is kept secret.
- The key session *K* is kept secret.
- The MAC key K_{MAC} is kept secret.
- The starting variable SV is kept secret.
- The starting variable for the MAC SV_{MAC} is kept secret.
- Both of the devices are validating the Message Agreement property, which means that contents of the received messages correspond to the sent messages [35].
- Both of the devices are validating the Non-injective Synchronisation requirement; this property states that everything we intended to happen in the protocol description also happens in the trace [35].
- Both of the devices are alive, which means that, as expected, as intended communication partners, they have executed some events [35].

8.4 Attacker profiles

In order to set a threat model, we need to identify different types of attackers.
These profiles have been inspired and modified from [16]:

- Script Kiddies:
  Script Kiddies are generally 14 - 16 years old and still at school preferring to spend their free time working on computers rather than playing with friends.

  They have rudimentary knowledge of the mechanics of the system, but enough knowledge to cause damage.

Script kiddies use the tools that are already freely available on the internet and go to internet forums to swap information, gain experience and guidance from others.

They may be driven by curiosity, a desire to test their skills or assess a weakness they have heard about. They may also be motivated by money or kudos.

- Expert individuals:
  Broadly speaking, Expert individuals understand how the system works. They also understand how to use some hacking tools.

  They are highly capable, intelligent and able to write good applications themselves because they have a sound knowledge of common programming languages. They probably have, or could acquire, reasonably good jobs in the IT industry on the basis of their abilities.

  They may have a wide range of motivations, ranging from some venal motivation to industrial espionage, via security analysts.

- Organized crime:
  This profile describes a group of Expert individuals. These people are normally involved with organized crime syndicates, governments or militant wings of political parties. They have the same kind of abilities as that of Expert individuals, but are lot more dangerous due to their number and the help that the rest of the structure (the organization) can bring to them either financially or informationally. Their aim is to make money or damage other countries' or parties infrastructures.

8.5 Security analysis

We have identified different kinds of attacks which cover the threat model presented in Chapter 3.3, i.e. template modification attacks, mimic attacks, data modification or injection and Denial of Service.

- Template modification attacks:
  These attacks correspond to all modification of the template either by False Enrolment or by modification of the template where it is stored.

  To proceed to a False Enrolment, the malicious individual needs to have a fake identity card and be able to answer some questions on the genuine user. Modifying the template while it is saved on the SC cannot be done as the enrolment takes place in a trusted and secure place. The modification of the template in the SC is extremely difficult due to the tamper resistance of this device.

- Mimic attack:
  This refers to any form of attempting to reproduce the biometric, here the gesture. This can be done either with a fake biometric or by Reuse of Residuals. This last form can only be used on an untrusted terminal as it requires extraction of the last attempt recorded from the memory of the terminal.

  It is easy for an individual to copy a gesture they have seen, but as we have discussed in Chapter 6, a mimic attack made by a human has only around 11% chance of succeeding.

  Another form of mimic attack would be to use a robotized mannequin with the rig biometric geometry. The mannequin has to be programmed to reproduce a pre-recorded gesture, such as residuals extracted from the memory of a terminal.
  This method has several drawbacks: this kind of mannequin is expensive and cumbersome, and is difficult to adapt to attack several genuine users.

- Data modification or injection:
  This represents several threat vectors:
  Replay Attacks / False Data Inject, Storage Channel Intercept and Data Inject, Match Override / False Match, Decision Override / False Accept, Fake Digital Biometric, Synthesised Feature Vector and Template reconstruction.

  All these above vectors attempt to break the communication between the SC and the terminal in order inject, modify or replay some data.

  - Replay Attacks / False Data Inject:
    This attack cuts the message flow between the SC and terminal and injects some malicious data.
    This data would be ignored by both SC and terminal unless this data includes the correct use of Id_{SC}, Id_T, and nonce. As Id_{SC} and Id_T do not change, the nonce is providing the protection against this attack.
  - Storage Channel Intercept and Data Inject:
    The storage element is the SC. It is very difficult for a malicious individual to access to the memory and modify it.
  - Match Override / False Match, Decision Override / False Accept:
    The protection against such attacks is mainly the fact that the decision message is encrypted, which means that even a little modification on this encrypted message

will impact the whole of the decrypted message. In turn, this results in a modification of the nonce in the message, which would be no longer match the one used.

  - Fake Digital Biometric, Synthesised Feature Vector and Template reconstruction: to implement these attacks, the malicious individual needs to claim to be a certified terminal, which results in a masquerade attack. We have seen in 8.3.2 that the SC and the terminal authenticate each other before beginning any transaction.

- Denial of Service:
  This represents any attack which aims to block the authentication of a person. This can be done by either damaging the system, or performing Override Feature Extraction or System Parameter Override / Modification attacks.

  The Overriding of the Feature Extraction is not possible because we have set up in a trusted environment. Similarly, System Parameter Override/Modification is not possible due to this trusted environment.

  Devices such as depth sensors can be easily disrupted by any environmental or malicious interference which makes any denial of service attack easy. Depth cameras are based on visual frequency, thus any object which can block or disturb this kind of wave will affect the whole system.

- Other threat vectors:
  We did not discuss about Modify Access Rights and System Interconnections because these threat vectors are out of the scope of our studies, as we are focusing on the authentication and not access rights management and system interconnection.

  Also, we did not discuss the Latent Print Reactivation as the GR system does not need any contact with a device, which implies than there is no latent print left.
  Unlike the ink left by the pen for signature recognition, GR does not leave any trace after the gesture has been done (or at least none that can be traced today).

8.6 Comparison with attackers' profiles

The assumptions as they have been described in 8.2.2 extremely limited the abilities of the attackers and are not representative of the reality. In this section, we relax these assumptions.

Table 8.2 summarizes the options that each profile has to attack the authentication system. Due to their lack of knowledge and technical skill, the Script Kiddies would be limited to basic attacks such as the denial of service by damaging the devices or mimicking the gesture, but their chance of success would be rather limited.

Most of the Expert individuals would be able to do the same attacks as those made by Script Kiddies, plus due to their high level of knowledge and technical skill, they might be able to do some others kinds of attacks such as:

- False enrolment: some Expert individuals may have the contact or the ability to forge a fake ID card and be able to answer the questions asked by the authority agent.
- Fake digital biometric: some Expert individuals may be able to afford a robot, which is able to replicate a human gesture, and be able to set up it. But the problem still remains of getting the biometric required by the system and bringing all machinery in front of the terminal without damaging the machinery and without arousing suspicion.

Organized crime is able to perform the same attacks as the previous profiles but it is also able to break the SC and terminal security which then gives access to the SK_T and the SC certificate $SK_A\{PK_{SC}\}$. This access allows all the attacks we have described above.

- The ability to access the template makes the Unauthorised template modification possible.
- A basic replay attack would still be impossible because of the protocol and the use of nonce, but a false data injection is possible due to knowledge of the terminal private key with which it is possible to decrypt the message where the session key is exchanged.
- With the access to both the template and private key, it would be easy for a malicious individual to extract the template from the SC and inject this data to get authenticated.
- With the private key it is possible to know the session key which makes it possible to override the feature extracted, as well as the decision to bypass the authentication. It is also possible to Override Feature Extraction and System parameter override/modification to produce a denial of service.

Table 8.2: *Attackers' profile ability*
where ✗*: this profile cannot do this kind of attack;*
✓*: this profile can do this kind of attack;*
–*: this attack is not applicable to our model*
(source: Benoit Ducray)

Attacks	Threat vectors	Script Kiddies	Expert individuals	Organized crime
Template modification	Unauthorised template modification	✗	✗	✗
	False enrolment	✗	✓	✓
Mimic attack	Fake physical biometric	✓	✓	✓
	Reuse of residuals	✗	✗	✓
Data modification or injection	Replay attacks/false data inject	✗	✗	✓
	Storage channel intercept and data inject	✗	✗	✓
	Override feature extraction	✗	✗	✓
	Decision override/false accept	✗	✗	✓
	Fake digital biometric	✗	✗	✓
	Synthesised feature vector	✗	✗	✓
	Template reconstruction	✗	✗	✓
Denial of service	Damaging the system	✓	✓	✓
	Override Feature Extraction	✗	✗	✓
	System parameter override/modification	✗	✗	✓
–	Modify access rights	–	–	–
	System interconnections	–	–	–
	Latent print reactivation	–	–	–

- With the terminal private key and the capture of its certificate (which is sent in clear in the protocol) it is possible to emulate a fake terminal which makes Synthesised feature vector and Template reconstruction possible.

Most of the profiles do not have more possibilities of bypassing the authentication other than just trying to mimic the gesture, which has only an 11% chance of getting accepted. Some Experts would be able to attempt a false enrolment by having knowledge of their victim's life and obtaining a fake identity card.

A group of these Experts would be able to bypass the SC and terminal protection, which leaves the system open to any form of attack. But attackers with such a profile would probably use the easiest way to achieve their goals rather than by breaking such strong protection.

8.7 Summary

We have seen in the previous chapter that an authentication based on GR may correctly authenticate the genuine user and keep away any malicious individual. We have also seen, that a gesture template and GR application can be stored on a SC.

However, all of this would be useless if the communication between the SC and the terminal is not secure. In this chapter we have proposed a security protocol which would be used between the SC and the terminal.

This protocol requires the involvement of an authority who certifies the authenticity of both the SC and the terminal, which implies both of these elements have a certificate signed by the authority.

The communication between the SC and terminal will begin by exchanging the certificate and the generation of two AES keys of 192 bytes generated by the SC.

One key will be used for encrypting the message (here the gesture) and the other to generate the MAC. Then the terminal will send the gesture attempt to the SC, which will compare it to the template on the SC. Then the SC will send the acceptance or rejection of the attempt.

The use of mutual authentication of the SC and terminal, the encryption of messages, and the repeated generation of nonces provide a protection against a wide range of attacks.

After describing some attackers' profiles, such as Script Kiddies, Expert individuals and Organized crime, we assess the protocol to different kind of attacks, ranging from fake enrolment to Override feature extraction via the use of fake biometrics.

Then we assessed the abilities of each profiles presented earlier to produce attack against the system. Most of the attacker profiles have limited options to attack, but the grouping of several Experts may be able to break all of the system's securities.

Part III

Conclusion

Chapter 9

Conclusion and Future Work

Contents

This chapter concludes the thesis by summarising the contributions and discusses potential future work.

9.1 Conclusion

The main goal of this thesis were:

- To explore the possible techniques to authenticate a person using changeable multi-factor authentication measures that are influenced by biometrics.
- To evaluate their security characteristics.
- And finally to show through an example, i.e authentication by Gesture Recognition (GR), that an authentication application using this kind of system can be secure.

We began the discussion by an overview of the three factors of authentication formalised in the sixties by IBM [112] (i.e. knowledge factors, ownership factors, inherence factors).
We then focused on biometric authentication and gave some general principles, the components and processes involved as well as an overview of the techniques and modalities.
We continued the discussion by presenting some background about GR, and the systems we used: sensors, features extraction and analyser.

After a review of some major security issues with ‘fixed’ biometrics and the limited solutions that exist, we defined a new family of biometrics, viz. dynamic biometrics.
This family gathers biometrics that are dynamic when physical / behavioural (inherent) biometric information is captured together with a knowledge factor from a user, such that it can be used as the basis of a one-step two-factor authentication.

We then investigated how the security of dynamic biometrics compares with conventional biometrics.
Evaluation criteria devised by Bonneau et al. [23] were used as a basis to assess the security of several fixed/dynamic biometrics. We found than the inclusion of a knowledge factor in a dynamic biometric brings some privacy benefits in comparison to ‘fixed’ biometrics. In addition to making the biometric changeable, dynamic biometrics improve Unlinkability.

With such information, we studied the ability of GR to be an efficient authenticator system.
We presented four possible attacks against GR (viz. “Brute Force attack”, “Dictionary attack”, “Storage Leakage” and “Shoulder Surfing attack”) and we set up two experiments in order to explore two out of the four attacks (i.e. “Brute Force attack” and “Shoulder Surfing attack”).
These experiments were done with different sensors: the Kinect$^{\text{TM}}$, which aims to track full body movement; and the Leap Motion which aims to track movement of the hands.

In the case of the experiment of a "Brute Force attack" using the Kinect$^{\text{TM}}$ system, none of the attacks succeeded despite 56 minutes of attack (representing over 10,000 attacking gestures). This result is at least as good as an unknown 4-digit PIN. In the case of a "Shoulder Surfing attack", the system has 1.7% FAR for 93% of TAR and an EER of 2.8%.

The Leap Motion experiment was focused on a "Shoulder Surfing attack" as all participants performed predetermined gestures. An attacker mimicking a known gesture had an 11.88% likelihood of a successful attack, whilst a genuine user had a 88.12% chance to be correctly authenticated.

We show that all these results were in line with other research done in this domain and using the same kind of sensor.

Then we made a comparison between fingerprint recognition and GR.
We established that for identifying or authenticating individuals, fingerprint is much more reliable than GR.

However, against attacks, GR is a better choice for rejecting an attacker unless we know the method used to forge the biometrics. Should the forging technique be known, the liveness detection algorithms for fingerprints obtain a better score but in the real world such knowledge is greatly unlikely.

One of biometrics systems' principal challenges is to store the biometric reference template in a secure location. We investigated this problem by implementing an SC and a smart phone HCE GR application in order to asses the feasibility of such applications on these devices with the DTW algorithm to compare gestures. We measured communication and processing time for both SC and HCE.

Although it is possible to run a GR application on an SC at platform level, the implementation took around a minute which is too long for an real application.
There are different ways to improve this time, for example running the application at a lower level either in hardware or with native code.

The HCE application was far more practical for a real world application, but does not provide the attack resistance that an SC offers. The use of a Trusted Execution Environment (TEE) may enhance security and resist logical and malware attacks, although a TEE does not offer the tamper-resistance of an SC.

Whatever the efficiency of an authentication system and how well-protected the template is, if the communication between the terminal and the SC is not secure any malicious individual can spy, inject or modify the communication. In order to secure this communication, we proposed a security protocol which could be used between the SC and the terminal. This protocol involves an authority who certifies the authenticity of both the SC and the terminal which implies both of these elements have a certificate signed by the authority.

The communication between the SC and the terminal starts with the exchange of the certificate and the generation of two AES keys of 192 bytes generated by the SC. One of these is used to encrypt the message (here, the gesture) and the other is used to generate the MAC. Then the terminal can start to send the gesture attempt to the SC, which will then do the comparison with the stored template. Once the whole gesture has been sent, the SC will send the acceptance or rejection of the attempt.

A mutual authentication of the SC and terminal, the encryption of messages, and the repeated generation of nonces protect against a wide range of attacks.

We assessed the protocol against different kinds of attack, ranging from fake enrolment to Override feature extraction via use of fake biometrics.

We also described some attackers' profiles as Script Kiddies, Expert Individuals and Organized crime and assess the abilities of each of these profiles to produce an attack against the system. It resulted that most of the attackers' profiles have limited options to attack but the grouping of several Experts may be able to break all of the system's security.

Dynamic biometrics, such as authentication based on GR, provide several advantages compared to 'fixed' biometrics, for example, the possibility to change the secret (which in the case of GR is the gesture), which improves Unlinkability. But dynamic biometric cannot be used to identify somebody, only for authentication; because of this, the FAR is higher than other biometric families.

Dynamic biometrics do not aim to replace 'traditional' biometrics.

Dynamic biometrics and 'fixed' biometrics can work together in order to identify and authenticate a user. A 'fixed' biometric can be used as a login, in much the same way as e-mail addresses are nowadays used as login-ID, with a dynamic biometric used to authenticate the identity claim. This operation would allow fixed biometric information to be shared widely with no more security problems than the use of e-mail addresses; the use of dynamic biometrics acts as a highly secure "password" because of the addition of the security properties of dynamic biometrics.

The collection of ‘fixed’ biometric can be transparent to the user, as it will be collected while providing the secret dynamic biometric. For example during the execution of a secret hand gesture, authentication authorities may as well collect fingerprints and/or the face of the user for identification.

9.2 Future work

Future work in the areas of dynamic biometrics and GR is now identified.

9.2.1 Dynamic biometric

It would be possible to develop completely new authentication methods based on the above definition of dynamic biometrics whether these would be feasible for practical use or not.
For example, in Chapter 5 we mentioned a Gesture Recognition (GR) based on electromyography, and also one based on Speaker Recognition.

It might be possible to create an authentication method based on a mix of these two dynamic biometrics with an electromyography of the mouth.
This technique would fit the dynamic biometrics definition as the movement of the lips depend on the text being spoken at the time of capture, and it can be easily changed.
The biometric component is the electromyograph.

The definition for each class of dynamic biometrics needs more precision as some of the dynamic biometrics may fall into different categories.
For example, GR based on electromyography can be considered to belong to the class of gesture-based as does GR, but it can also be classified as thought-based as the electromyography can be associated to thought.

Here we can base the classification either on the information that is captured, i.e. the data which may be visual information or electric impulses, or on what is required from the user.

9.2.2 Gesture recognition

As we have seen the Chapter 7, running a GR authentication process on an SC takes a lot of time. Our implementation has been done at the application level but an implementation in the hardware and/or the native SC code may reduce the process time to a third of the time found in our experiment.

New solutions should be explored, in a context of improved security and accuracy of the whole process, taking into account, on the one hand, the continuing improvement of calculation, processing and storage capabilities (Moore's Law), and on the other hand, the bare minimum data (number of frames, features and the relationship between those two) that would be needed to fulfil the following:

- Improve the time required by the authentication process,
- Reduce the size of the necessary storage memory.

A problem we did not tackle in this research, relates to the evolution of the way a gesture is performed by a given individual over time, especially when the gesture is performed at varying intervals. Habits and physical factors might play a role here.
For this problem, Liu et al. [72] propose the solution of storing two templates of the reference gesture and of updating one of them at each correct authentication in order to keep up with the gesture's evolution.
However, this solution may result in a template being updated with attackers details following a successful attack. Thus, an alternative, secure method for tracking the evolution of a gesture should be studied.

Several experiments can be imagined where both 'fixed' and dynamic biometrics would be used together for authentication. An example of an application which uses both biometrics would be one that used the 'fixed' biometrics as a login the way an e-mail address works, and the dynamic biometric as the equivalent of a password.

Bibliography

[1] ISO 7816-4. http://www.cardwerk.com/smartcards/smartcard_standard_ISO7816-4.aspx, 2013. [Online; accessed 11 November 2016]. 104

[2] Bionym nymi. http://www.bionym.com/, 2016. [Online; accessed 23 May 2016]. 73

[3] Leap Motion. https://developer.leapmotion.com/getting-started/javascript/developer-guide, 2016. [Online; accessed 25 February 2016]. 64, 99

[4] Leap Motion. https://developer.leapmotion.com/documentation/cpp/unreal/Leap_Unreal_Cpp_Tutorial.html, 2016. [Online; accessed 16 November 2016]. 82

[5] Structured-light 3D scanner. https://kinect1.wordpress.com/future/, 2016. [Online; accessed 01 December 2016]. 62

[6] Japan researchers warn of fingerprint theft from 'peace' sign. https://phys.org/news/2017-01-japan-fingerprint-theft-peace.html, 2017. [Online; accessed 27 April 2017]. 71

[7] G. Aggarwal, N. K. Ratha, T.-Y. Jea, and R. M. Bolle. Gradient based textural characterization of fingerprints. In *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on*, pages 1–5. IEEE, 2008. 38

[8] Z. Ahmad, K. E. Mayes, S. Dong, and K. Markantonakis. Considerations for mobile authentication in the cloud. *information security technical report*, 16(3):123–130, 2011. 49

[9] Z. Akhtar and N. Alfarid. Robustness of serial and parallel biometric fusion against spoof attacks. *Computer Networks and Intelligent Computing; Venugopal, KR, Patnaik, LM, Eds*, pages 217–225, 2011. 35

[10] Z. Akhtar and S. Kale. Security analysis of multimodal biometric systems against spoof attacks. In *International Conference on Advances in Computing and Communications*, pages 604–611. Springer, 2011. 35

[11] F. Alshanketi, I. Traore, and A. A. Ahmed. Improving Performance and Usability in Mobile Keystroke Dynamic Biometric Authentication. In *Security and Privacy Workshops (SPW), 2016 IEEE*, pages 66–73. IEEE, 2016. 73

[12] R. Ang, R. Safavi-Naini, and L. McAven. Cancelable key-based fingerprint templates. In *Information Security and Privacy*, pages 242–252. Springer, 2005. 52, 71

[13] Y. Apelbaum. *User Authentication Principles, Theory and Practice*. Fuji Technology Press, 2007. 20, 22

[14] I. Aslan, A. Uhl, A. Meschtscherjakov, and M. Tscheligi. Mid-air authentication gestures: an exploration of authentication based on palm and finger motions. In *Proceedings of the 16th International Conference on Multimodal Interaction*, pages 311–318. ACM, 2014. 93

[15] M. T. I. Aumi and S. Kratz. Airauth: evaluating in-air hand gestures for authentication. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services*, pages 309–318. ACM, 2014. 70, 93, 94

[16] R. Barber. Hackers profiled-who are they and what are their motivations? *Computer Fraud & Security*, 2001(2):14–17, 2001. 114

[17] E. B. Barker and J. M. Kelsey. Sp 800-90a. recommendation for random number generation using deterministic random bit generators. 2012. 113

[18] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on neural networks*, 13(6):1450–1464, 2002. 36

[19] G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural computation*, 12(10):2385–2404, 2000. 36

[20] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997. 36

[21] L. Biel, O. Pettersson, L. Philipson, and P. Wide. ECG analysis: a new approach in human identification. *IEEE Transactions on Instrumentation and Measurement*, 50(3):808–812, 2001. 45

[22] D. Blackburn, C. Miles, and B. Wing. Biometrics" foundation documents". Technical report, DTIC Document, 2009. 29, 30, 36, 37, 39, 41, 42, 44, 45

[23] J. Bonneau, C. Herley, P. C. Van Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *2012 IEEE Symposium on Security and Privacy*, pages 553–567. IEEE, 2012. 16, 70, 74, 79, 124

[24] J. Brainard, A. Juels, R. L. Rivest, M. Szydlo, and M. Yung. Fourth-factor authentication: somebody you know. In *Proceedings of the 13th ACM conference on Computer and communications security*, pages 168–178. ACM, 2006. 22

[25] C. Cadoz. Les réalités virtuelles. 1994. 59

[26] A. Chahar, S. Yadav, I. Nigam, R. Singh, and M. Vatsa. A leap password based verification system. In *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*, pages 1–6. IEEE, 2015. 93

[27] J. Chuang. One-Step Two-Factor Authentication with Wearable Bio-Sensors. 70, 73

[28] J. Chuang, H. Nguyen, C. Wang, and B. Johnson. I think, therefore I am: Usability and security of authentication using brainwaves. In *International Conference on Financial Cryptography and Data Security*, pages 1–16. Springer, 2013. 73

[29] G. D. Clark and J. Lindqvist. Engineering gesture-based authentication systems. *IEEE Pervasive Computing*, 14(1):18–25, 2015. 83

[30] P. Cofta, S. Furnell, and H. Lacohee. *Understanding Public Perceptions: Trust and engagement in ICT-mediated services*. Intl. Engineering Consortiu, 2008. 70, 78

[31] P. Coli, G. L. Marcialis, and F. Roli. Power spectrum-based fingerprint vitality detection. In *2007 IEEE Workshop on Automatic Identification Advanced Technologies*, 2007. 72

[32] W. B. Committee et al. *Biometric recognition: challenges and opportunities*. National Academies Press, 2010. 32, 54, 56

[33] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995. 36

[34] F. F. I. E. Council. FFIEC Releases Supplemental Guidance on Internet Banking Authentication. *DC: Federal Deposit Insurance Corp.(FDIC). Retrieved June*, 28:2011, 2011. 22, 23, 25, 26, 27

[35] C. Cremers and S. Mauw. *Operational semantics and verification of security protocols*. Springer Science & Business Media, 2012. 114

[36] J. Cross and C. Smith. Thermographic imaging of the subcutaneous vascular network of the back of the hand for biometric identification. In *Security Technology, 1995. Proceedings. Institute of Electrical and Electronics Engineers 29th Annual 1995 International Carnahan Conference on*, pages 20–35. IEEE, 1995. 45

[37] D. Damopoulos, G. Kambourakis, and S. Gritzalis. From keyloggers to touchloggers: Take the rough with the smooth. *Computers & security*, 32:102–114, 2013. 73

[38] T. Darrell and A. Pentland. Space-time gestures. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, pages 335–340. IEEE, 1993. 65

[39] J. Daugman. How iris recognition works. *IEEE Transactions on circuits and systems for video technology*, 14(1):21–30, 2004. 41

[40] B. Ducray, S. Cobourne, K. Mayes, and K. Markantonakis. Authentication based on a changeable biometric using gesture recognition with the kinect^{TM}. In *2015 International Conference on Biometrics (ICB)*, pages 38–45. IEEE, 2015. 70, 77

[41] B. Ducray, S. Cobourne, K. Mayes, and K. Markantonakis. Gesture recognition implemented on a personal limited device. In *Information and Communication Systems (ICICS), 2017 8th International Conference on*, pages 171–176. IEEE, 2017. 97

[42] L. Ducray. *Contributing to a better understanding of the 'international-multicultural' tension in a team : 4 exploratory cases of IS implementation projects ERP type*. PhD thesis, Universite Paris Est, 2013. 59

[43] T. Dunstone and N. Yager. *Biometric system and data analysis: Design, evaluation, and data mining*. Springer Science & Business Media, 2008. 30, 31, 33, 34, 53, 55

[44] K. Eagles, K. Markantonakis, and K. Mayes. A comparative analysis of common threats, vulnerabilities, attacks and countermeasures within smart card and wireless sensor network node technologies. In *Information Security Theory and Practices. Smart Cards, Mobile and Ubiquitous Computing Systems*, pages 161–174. Springer, 2007. 23, 24

[45] J.-E. Ekberg, K. Kostiainen, and N. Asokan. Trusted execution environments on mobile devices. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 1497–1498. ACM, 2013. 104

[46] D. Gavrila and L. Davis. Towards 3-D model-based tracking and recognition of human movement: a multi-view approach. In *International workshop on automatic face-and gesture-recognition*, pages 272–277. Citeseer, 1995. 65

[47] L. Ghiani, P. Denti, and G. L. Marcialis. Experimental results on fingerprint liveness detection. In *Articulated Motion and Deformable Objects*, pages 210–218. Springer, 2012. 95

[48] P. Gloor. Bertillon's method and anthropological research; a new use for old anthropometric files. *Journal of the Forensic Science Society*, 20(2):99–101, 1980. 40

[49] C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, J. T. McConville, I. Tebbetts, and R. A. Walker. Anthropometric survey of us army personnel: Summary statistics, interim report for 1988. Technical report, DTIC Document, 1989. 83

[50] T. Guardian. US government hack stole fingerprints of 5.6 million federal employees. `https://www.theguardian.com/technology/2015/sep/23/us-government-hack-stole-fingerprints`, 2015. [Online; accessed 13 December 2016]. 71

[51] E. T. Hall, R. L. Birdwhistell, B. Bock, P. Bohannan, A. R. Diebold Jr, M. Durbin, M. S. Edmonson, J. Fischer, D. Hymes, S. T. Kimball, et al. Proxemics [and comments and replies]. *Current anthropology*, pages 83–108, 1968. 59

[52] O. Henniger and K. Franke. Biometric user authentication on smart cards by means of handwritten signatures. In *Biometric Authentication*, pages 547–554. Springer, 2004. 104

[53] R. Heyer. Biometrics technology review 2008. 2008. 70, 71, 94

[54] M. S. Holi. Electromyography analysis for person identification. *International Journal of Biometrics and Bioinformatics (IJBB)*, 5(3):172, 2011. 73

[55] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold. ECG to identify individuals. *Pattern recognition*, 38(1):133–142, 2005. 45

[56] A. Jain, A. Ross, and S. Prabhakar. Fingerprint matching using minutiae and texture features. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 3, pages 282–285. IEEE, 2001. 38

[57] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle. An identity-authentication system using fingerprints. *Proceedings of the IEEE*, 85(9):1365–1388, 1997. 29

[58] A. K. Jain, K. Nandakumar, X. Lu, and U. Park. Integrating faces, fingerprints, and soft biometric traits for user recognition. In *International Workshop on Biometric Authentication*, pages 259–269. Springer, 2004. 38

[59] A. K. Jain, A. Ross, and S. Pankanti. Biometrics: a tool for information security. *IEEE transactions on information forensics and security*, 1(2):125–143, 2006. 36, 39, 40, 41, 42, 50, 52

[60] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004. 30, 77, 78, 94

[61] B. Johnson, T. Maillart, and J. Chuang. My thoughts are not your thoughts. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 1329–1338. ACM, 2014. 73, 78

[62] G. Kambourakis, D. Damopoulos, D. Papamartzivanos, and E. Pavlidakis. Introducing touchstroke: keystroke-based authentication system for smartphones. *Security and Communication Networks*, 2014. 73

[63] E. Keogh and C. A. Ratanamahatana. Exact indexing of Dynamic Time Warping. *Knowledge and information systems*, 7(3):358–386, 2005. 66, 98

[64] Z. Kleinman. Politician's fingerprint 'cloned from photos' by hacker. `http://www.bbc.co.uk/news/technology-30623611`, 2014. [Online; accessed 13 December 2016]. 71

[65] L. Kratz, D. Morris, and T. S. Saponas. Making gestural input from arm-worn inertial sensors more practical. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1747–1750. ACM, 2012. 61

[66] A. Kumar, D. C. Wong, H. C. Shen, and A. K. Jain. Personal authentication using hand images. *Pattern Recognition Letters*, 27(13):1478–1486, 2006. 39, 40

[67] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113, 1997. 36

[68] X. Leng. Smart card applications and security. *information security technical report*, 14(2):36–45, 2009. 23, 24

[69] J. F. Lichtenauer, E. A. Hendriks, and M. Reinders. Sign language recognition by combining statistical DTW and independent classification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11):2040–2046, 2008. 65

[70] lita Person. *World 3D Camera Market - Opportunities and Forecasts, 2013 - 2020*. Jan 2015. 81

[71] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan. uWave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5(6):657–675, 2009. 70

[72] J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan. uwave: Accelerometer-based personalized gesture recognition and its applications. *Pervasive and Mobile Computing*, 5(6):657–675, 2009. 128

[73] T. O. Majekodunmi and F. E. Idachaba. A review of the fingerprint, speaker recognition, face recognition and iris recognition based biometric identification technologies. 2011. 74

[74] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar. *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009. 38, 72

[75] E. Marasco and A. Ross. A survey on antispoofing schemes for fingerprint recognition systems. *ACM Computing Surveys (CSUR)*, 47(2):28, 2015. 77

[76] G. L. Marcialis, F. Roli, and A. Tidu. Analysis of fingerprint pores for vitality detection. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1289–1292. IEEE, 2010. 72

[77] K. Mayes, F. Piper, and K. Markantonakis. Smart card based authentication: any future. *Computers & Security*, 24:188–191, 2005. 20, 24, 29, 51, 52

[78] Microsoft. PrimeSense Supplies 3-D-Sensing Technology to "Project Natal" for Xbox 360. `http://www.microsoft.com/en-us/news/press/2010/mar10/03-31primesensepr.aspx`, 2010. [Online; accessed 31 March 2010]. 64

[79] Microsoft. Human Interface Guidelines v1.8. `http://go.microsoft.com/fwlink/?LinkID=247735`, 2013. [Online; accessed 09 September 2013]. 64, 82

[80] Microsoft. Kinect for Windows features. `http://www.microsoft.com/en-us/kinectforwindows/discover/features.aspx`, 2013. [Online; accessed 17 September 2013]. 64

[81] Microsoft. Kinect for Windows Sensor Components and Specifications. `http://msdn.microsoft.com/en-us/library/jj131033.aspx`, 2014. [Online; accessed 27 January 2014]. 63, 82

[82] A. Mir, S. Rubab, and Z. Jhat. Biometrics verification: a literature survey. *International Journal of Computing and ICT Research*, 5(2):67–80, 2011. 26, 29, 30, 36, 37, 42, 45

[83] S. Mitra and T. Acharya. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(3):311–324, 2007. 61, 62, 63

[84] S. Nikam and S. Agarwal. Fingerprint liveness detection using curvelet energy and co-occurrence signatures. In *Computer Graphics, Imaging and Visualisation, 2008. CGIV'08. Fifth International Conference on*, pages 217–222. IEEE, 2008. 72

[85] S. B. Nikam and S. Agarwal. Texture and wavelet-based spoof fingerprint detection for fingerprint biometric systems. In *Emerging Trends in Engineering and Technology, 2008. ICETET'08. First International Conference on*, pages 675–680. IEEE, 2008. 72

[86] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003. 77

[87] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002. 72

[88] P. S. Penev and J. J. Atick. Local feature analysis: A general statistical theory for object representation. *Network: computation in neural systems*, 7(3):477–500, 1996. 36

[89] N. Prakash. Host card emulation. *International Journal of Scientific and Research Publications*, 5(8):1–3, 2015. 98, 99

[90] J. Putz-Leszczyńska and M. Kudelski. Hidden signature for DTW signature verification in authorizing payment transactions. *Journal of telecommunications and information technology*, pages 59–67, 2010. 102

[91] N. K. Ratha, S. Chikkerur, J. H. Connell, and R. M. Bolle. Generating cancelable fingerprint templates. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):561–572, 2007. 52, 71

[92] C. Rathgeb, F. Breitinger, and C. Busch. Alignment-free cancelable iris biometric templates based on adaptive bloom filters. In *2013 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2013. 52, 71

[93] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978. 108

[94] C. Roberts. Biometric attack vectors and defences. *Computers & Security*, 26(1):14–25, 2007. 16, 46, 47

[95] R. N. Rodrigues, N. Kamat, and V. Govindaraju. Evaluation of biometric spoofing in a multimodal system. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–5. IEEE, 2010. 35

[96] A. Ross and A. K. Jain. Multimodal biometrics: An overview. In *Signal Processing Conference, 2004 12th European*, pages 1221–1224. IEEE, 2004. 33, 34

[97] S. K. Sahoo, T. Choubisa, and S. M. Prasanna. Multimodal biometric person authentication: A review. *IETE Technical Review*, 29(1):54–75, 2012. 26

[98] M. Sandström. Liveness detection in fingerprint recognition systems. 2004. 72

[99] N. SCIENCE and T. COUNCIL. The national biometrics challenge. In *National Science and Technology Council Subcommittee on Biometrics and Identity Management*, pages 1–46. EXECUTIVE OFFICE OF THE PRESIDENT NATIONAL SCIENCE AND TECHNOLOGY COUNCIL, 2011. 49, 50, 55, 56

[100] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013. 64

[101] S. J. Simske. Dynamic biometrics: The case for a real-time solution to the problem of access control, privacy and security. In *2009 First IEEE International Conference on Biometrics, Identity and Security (BIdS)*, pages 1–10. IEEE, 2009. 72

[102] C. Sousedik and C. Busch. Presentation attack detection methods for fingerprint recognition systems: a survey. *IET Biometrics*, 3(4):219–233, 2014. 72, 95

[103] A. Squicciarini and E. Bertino. Privacy preserving multi-factor authentication with biometrics. 2006. 20, 29, 30, 32, 51, 52, 53

[104] B. Tan and S. Schuckers. Liveness detection for fingerprint scanners based on the statistics of wavelet signal processing. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 26–26. IEEE, 2006. 72

[105] B. Tan and S. Schuckers. New approach for liveness detection in fingerprint scanners based on valley noise analysis. *Journal of Electronic Imaging*, 17(1):011009–011009, 2008. 72

[106] G. Ten Holt, M. Reinders, and E. Hendriks. Multi-dimensional dynamic time warping for gesture recognition. In *Thirteenth annual conference of the Advanced School for Computing and Imaging*, volume 300, 2007. 65

[107] J. Thorpe, P. C. van Oorschot, and A. Somayaji. Pass-thoughts: authenticating with our minds. In *Proceedings of the 2005 workshop on New security paradigms*, pages 45–56. ACM, 2005. 73

[108] J. Tian, C. Qu, W. Xu, and S. Wang. Kinwrite: Handwriting-based authentication using kinect. In *NDSS*, 2013. 93, 94

[109] R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, S. Ricerche, and F. Roli. Fusion of multiple clues for photo-attack detection in face recognition systems. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–6. IEEE, 2011. 77

[110] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991. 36

[111] A. Umar, K. Mayes, and K. Markantonakis. Performance variation in host-based card emulation compared to a hardware security element. In *Mobile and Secure Services (MOBISECSERV), 2015 First Conference on*, pages 1–6. IEEE, 2015. 99

[112] E. L. Van Den Broek. Beyond biometrics. *Procedia Computer Science*, 1(1):2511–2519, 2010. 15, 22, 29, 43, 53, 55, 124

[113] V. Velichko and N. Zagoruyko. Automatic recognition of 200 words. *International Journal of Man-Machine Studies*, 2(3):223–234, 1970. 65

[114] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler. Analysis of the accuracy and robustness of the leap motion controller. *Sensors*, 13(5):6380–6393, 2013. 64

[115] D. A. Winter. *Biomechanics and motor control of human gait: normal, elderly and pathological*. 1991. 27, 39, 40, 42, 43, 45

[116] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. Von Der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):775–779, 1997. 36, 38

[117] J. Wu, J. Konrad, and P. Ishwar. Dynamic Time Warping for gesture-based user identification and authentication with kinect. In *Acoustics, Speech and Signal Processing*

(ICASSP), 2013 IEEE International Conference on, pages 2371–2375. IEEE, 2013. 93, 94

[118] A. L. Yuille. Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991. 36

Appendix A

Scyther code

```
usertype Gesture;
    usertype Accept;
    secret Cert: Function;

    protocol ExampleProtocol(T,SC) {

    role T {

    fresh Nt1: Nonce;
    fresh Nt2: Nonce;
    const gesture1: Gesture;

    var K: Nonce;
    var Kmac: Nonce;
    var SV: Nonce;
    var SVmac: Nonce;
    var Nsc1: Nonce;
    var Nsc2: Nonce;
    var accept: Accept;

    send˙1(T,SC, Cert(T), T, Nt1);
    recv˙2(SC,T, {K, Kmac, SV, SVmac, Cert(SC), SC, T, SVmac, Nsc1, Nt1}pk(T));
    send˙3(T, SC, {gesture1, T, SC, Nt2pk(SC), Nsc1}K, {{gesture1, T, SC, Nt2pk(SC), Nsc1}K}Kmac);
    recv˙4(SC,T, {accept, T, SC, Nt2, Nsc2}K, {{accept, T, SC, Nt2, Nsc2}K}Kmac);
```

```
claim˙T1(T, Secret, gesture1);

claim˙T2(T, SKR, K);
claim˙T3(T, SKR, Kmac);
claim˙T4(T, SKR, SV);
claim˙T5(T, SKR, SVmac);

claim˙T6(T, Niagree);
claim˙T7(T, Nisynch);
claim˙T8(T, Alive);

}

role SC{

fresh K: Nonce;
fresh Kmac: Nonce;
fresh Nsc1: Nonce;
fresh Nsc2: Nonce;
fresh SV: Nonce;
fresh SVmac: Nonce;
const accept: Accept;

var Nt1: Nonce;
var Nt2: Nonce;
var gesture1: Gesture;

recv˙1(T,SC, Cert(T), T, Nt1);
send˙2(SC,T, {K, Kmac, SV, SVmac, Cert(SC), SC, T, SVmac, Nsc1, Nt1}pk(T));
recv˙3(T, SC, {gesture1, T, SC, {Nt2}pk(SC), Nsc1}K, {{gesture1, T, SC, {Nt2}pk(SC), Nsc1}K}Kmac);
send˙4(SC,T, {accept, T, SC, Nt2, Nsc2}K, {{accept, T, SC, Nt2, Nsc2}K}Kmac);

claim˙SC1(SC, Secret, gesture1);

claim˙SC2(SC, SKR, K);
claim˙SC3(SC, SKR, Kmac);
```

```
    claim˙SC4(SC, SKR, SV);
    claim˙SC5(SC, SKR, SVmac);

    claim˙SC6(SC, Niagree);
    claim˙SC7(SC, Nisynch);
    claim˙SC8(SC, Alive);
}
};
```

Appendix B

Smart Card and Host Card Emulation times measured extended (Times measured in millisecond)

	Size of the APDU	Number of APDU sent	Communication time per APDU		Standard deviation of the communication time		Average time for the full application		Standard deviation of the full application		Estimate processing time	
			SC	HCE	SC	HCE	SC	HCE	SC	HCE	SC	HCE
G1	36	90	6.71	23.65	0.26	1.75	92982.15	2228.48	413.70	44.98	92378.09	99.48
	216	15	24.72	77.54	0.22	12.09	76029.02	1196.35	1055.72	20.64	75658.16	33.12
G2	36	49	6.71	23.65	0.26	1.75	50622.60	1214.52	276.90	30.31	50293.73	55.40
	216	9	24.72	77.54	0.22	12.09	43760.30	667.18	2653.56	23.05	43555.80	23.14
G3	66	90	9.79	31.81	0.34	3.79	108674.40	3084.76	408.12	50.06	107792.70	221.60
	198	30	23.76	71.80	0.34	4.94	94083.19	2253.42	1038.14	28.95	93370.15	99.30
G4	66	49	9.79	31.81	0.34	3.79	32448.10	1684.36	138.99	47.00	31968.07	125.52
	198	17	23.76	71.80	0.34	4.94	28092.53	1237.19	137.83	25.72	27702.44	64.68