Harokopio University
Informatics and Telematics Department

Keystroke Dynamics and User Authentication

Thesis by

Avramidis N. Loukas

Submitted in partial fulfillment of the requirements

for the degree of Bachelor of Informatics and Telematics.

Thesis Committee:
Dimitrios Michail
Vassiliki Andronikou
Iraklis Varlamis

Athens, September 2012

Acknowledgments

I would very much like to thank all the people who helped me finish this research, including my teachers, family, friends and fellow students.

First of all, Dr. Vassiliki Andronikou, who helped me a lot with the structure of my thesis and also introduced me into the field of keystroke dynamics, which has fascinated me in so many ways in the past year that I have been studying it, and will probably be the basic subject in my future work.

My instructors, Dr. Dimitrios Michail and Dr. Iraklis Varlamis, for helping me with the implementation and the completion of my thesis, and also for being in the thesis committee for my bachelor degree, it is much appreciated.

My fellow students and friends, who helped me test the mechanism that I created, in order to determine its accuracy and functionality.

My teachers in the Harokopio University of Athens, who taught me so much about informatics and telematics, and have always helped me with anything I needed.

And last but not least, I want to thank my father, who has supported me throughout all these years that I have been studying away from home, so I could get my bachelor’s degree and have a better life.

Loukas Avramidis

September 2012

Table of Contents

Abstract ix

1 Introduction 1

1.1 Biometric Systems 1

1.2 Vision and Related Research 2

1.3 Thesis Contribution 3

1.4 Thesis Structure 4

2 Biometric Systems 6

2.1 Identity Management 6

2.2 Identification Characteristics 7

2.3 Historic Review of Biometrics-based Identification and Verification 7

2.4 Qualitative Characteristics of Biometrics 9

2.5 Biometric Features and Systems Categorization 10

2.6 Biometric Features: Examples 12

2.6.1 The Face 12

2.6.2 Fingerprints 14

2.6.3 Keystroke Dynamics 16

2.7 Biometrics: Applications 17

2.8 The two Operational Phases of Biometric Systems 18

2.9 Types of Errors in Biometric Systems 18

3 Keystroke Sensitive Password Authentication Mechanism 21

3.1 Statement of Problem 21

3.2 Methodology 23

3.3 Classification Algorithms 24

4 Mechanism Implementation 30

4.1 Programming Environmental 30

4.2 Training Data Collection 30

4.3 Classification 32

4.4 Run-time and Interface Screen-shots 34

5 Mechanism Evaluation 45

5.1 Testing Environment 45

5.2 Results 46

5.3 Conclusion 47

6 Summary 50

Glossary 52

References 54

Table of Figures

Figure 1 Means for Identification 7
Figure 2 Classifications of Biometric Features and Systems 12
Figure 3 Seven common patterns used by the FBI [18] 16
Figure 4 Latencies between keystrokes when writing the word “password” by three different people. [48] 17
Figure 5 Relationship between FAR, FRR, and EER. [47] 20
Figure 6 The down-up and up-down features of a password. 24
Figure 7 Contours of the decision surfaces for the Euclidean classifier. 25
Figure 8 Contours of the decision surfaces for the Manhattan classifier. 26
Figure 9 Contours of the decision surfaces for the Scaled Manhattan classifier. 28
Figure 10The Keystroke Trainer screen at Training Mode. 35
Figure 11 The “Keystroke Trainer” screen, after accepting a password sample. 36
Figure 12 The “Keystroke Trainer” screen, after a wrong password entry. 36
Figure 13 The “Keystroke Trainer” screen, upon completion of the training data collection. 37
Figure 14 The “Authentication” screen. 37
Figure 15 The “Authentication” screen, with all classification algorithms granting access. 38
Figure 16 The “Authentication” screen, with all classification algorithms denying access. 38
Figure 17 The “Authentication” screen, with only one classification algorithm granting access. 39
Figure 18 Sample of the “output” XML containing the raw keystroke data. 40
Figure 19 Sample of the “featureOutput” XML, containing the features extracted from the raw data. 41
Figure 20 Sample of the “euclideanStats.xml” file, containing the anomaly score mean vectors created by the Euclidean algorithm. 42
Figure 21 Sample of the “manhattanStats.xml” file, containing the anomaly score mean vectors created by the Manhattan algorithm. 43
Figure 22 Sample of the “scaledManhattanStats.xml” file, containing the anomaly score mean vectors and the standard deviations created by the Scaled Manhattan algorithm. 44

Table of Tables

Table 1 Different evaluation procedures used by studies. [29] 22
Table 2 Accuracy Results of the Mechanism in terms of FRR and FAR 47

Abstract

Keystroke dynamics comprises a biometric feature that has been around since people started using computer keyboards. It is a very useful, inexpensive and easily portable way to determine the originality of a user who is typing a password. Passwords are easy to be compromised, but if an authentication mechanism, such as keystroke dynamics, were to be implemented along with the password system, it would greatly increase the security of any system.

The implementation so ever is somewhat complicated. The complexity is great and the difference in typing rhythms can vary a lot, which makes it really hard to define a certain unique typing profile for each individual. Researchers all over the world are trying to increase the accuracy of those kinds of mechanisms, accuracy which is measured in the probability of the mechanism denying access to the rightful user, and the probability of the mechanism grating access to the uncertified users, which both have to as low as possible, preferably under 2% respectively.

That can only be achieved by understanding the influential factors, which depend on the methods that are used, and will determine the outcome of the experiment and hence, the quality of the mechanism that was developed. Towards this direction, this thesis presents the biometric systems advancements and limitations focusing on keystroke dynamics. Furthermore, it describes a series of keystrokes dynamics mechanisms developed which are based on a set of selected features for user authentication purposes.

1 Introduction

In this chapter biometric systems are introduced. Research conducted in the field of biometrics is briefly presented and an introduction to the thesis contribution and structure is provided.

1.1 Biometric Systems

Biometric technologies have been the subject of many debates in the past few years, while governments and companies alike show great interest in the use of technologies based on biometric features, mostly in the fields of authentication and identification, such as in the case of biometric passports. This interest has grown even more because of the continuous increase of the commercial use of biometric applications and their capabilities as well as the heavy research activity in the field of biometrics, which has drastically increased the trustworthiness of modern systems. The use of biometric features for the authentication or identification of a person does not comprise a novel approach, but has been around since the early years of humanity, and is based in the ability of a person to recognize and distinguish someone through some features innate to the person, such as their face, voice, or the way they walk.

The term *biometric* (*βιομετρικός)* comes the Greek words “*βίος*” which means life and “*μετρικός*” which means measure. Hence, biometric technologies are based on human characteristics with certain qualities as we will see in Chapter 3, that allow us, with the proper extraction, analysis and processing, to identify the person they belong to. According to Jain et al [1], “*a biometric system is essentially a pattern recognition system that operates by acquiring biometric data from an individual, extracting a feature set from the acquired data, and comparing this feature set against the template set in the database*”.

Bromme [2] , in his attempt to describe the importance of biometric systems in the field of

information security, mentions that they include an authentication method based on the collection of human biological characteristics with the aid of sensors, the extraction of biometric features (non-processed mathematical data), the calculation of the biometric data and the creation of biometric prototypes, and finally, the authentication and identification based on those prototypes through the application of mathematical algorithms.

The Biometric Consortium [3] defines biometric systems as automated methods to authenticate a person based on a physiological or behavioral characteristic. Examples of such characteristics are the face, the fingerprints, the hand geometry, the way someone writes, the iris and the voice.

1.2 Vision and Related Research

In the past few years, research in biometric systems has advanced rapidly all over the world, with a vast number of researchers heavily working on advancements in different fields such as technology, law, social sciences and economics. Especially in the technological area, specialists work in a variety of domains in order to advance or build new algorithms for increasing the speed, the accuracy and the reliability of those systems. At the same time, by taking into consideration the great variety of human characteristics, researchers look for new biometric features that can be used to increase the accuracy of identification/verification systems and to reduce the cost of their application and their intrusiveness in human life. For example, a rather new biometric feature has been researched upon the past years, the otoacoustic emissions (OAEs), which is based on sounds that the ear produces when it is stimulated. According to the Engineering and Physical Sciences Research Council (EPSRC) [4] those emissions are created by the ability of the human body to boost very small sounds and according to their researches, those emissions are unique for every person. They are produced by the spiral screw, an organ in the human ear.

At the same time, great efforts are being made to apply biometrics for the advancement of security systems. In consideration of the required processing power and the privacy and information security issues [5] [6] [7] [8] [9], which are bound with the use of biometric systems, the scientific interest has befallen on the use of the already existing tools and their advancement, so they can support the computing requirements of any biometric applications.

1.3 Thesis Contribution

In this research, we are going to introduce and analyze a very interesting biometric feature; the keystroke dynamics. According to Kevin S. Killourhy [15][15], keystroke dynamics is the technology to distinguish computer users by their typing rhythms. The same way every person has a unique handwriting pattern, they also have a unique keystroke pattern, which can be used to authenticate or identify someone.

Keystroke dynamics have been a subject of research for the past thirty years, but despite the huge effort, they are still not demonstrably reliable enough. If they were, however, they would have significantly advanced computer security. For criminal investigations, keystroke dynamics could tie a suspect to the “scene” of a computer-based crime much like a fingerprint does in real-world crime. For access control, keystroke dynamics could act as a second factor in authentication; an impostor who compromised a password would still need to type it with the correct rhythm. For insider-threat detection, keystroke dynamics could detect when a masquerader is using another user’s account; the technology could even identify who is using a back-door account [15].

Within this context, we are going to demonstrate how we can authenticate a person according to their keystroke pattern when typing a password. Initially, the system is trained by the user, by typing a certain password several times, so that his biometric features - on how he types the specific password - can be collected. After we collect a certain amount of password entries from the original user, certain distance algorithms are performed on the keystroke timings. Following this, other people will act as impostors, by trying to enter the system with the same password. The purpose is to demonstrate how keystroke dynamics can be applied to deny access to impostors, who will actually know the password. Accuracy results will be expressed in terms of false-alarms and miss rates. The false-alarm rate is the percentage of test samples from the original user that will be denied by the system; the miss rate is the percentage of test samples from other typists that will be accepted by the system [15]. Hence, the goal is to reduce those error rates as much as possible.

1.4 Thesis Structure

The structure of this thesis is the following:

Chapter 1 Introduction

Introduction of the term “biometric system” and the recent findings in the field are reviewed. The way biometric features can be implemented in security systems is also explained, and keystroke dynamics are presented as an authentication method similar to the handwriting signature.

Chapter 2 Biometric Systems

Information on biometric systems is presented. The importance of identity management and various features that can be used for biometric identification and authentication are provided. The history of biometrics is briefly reviewed and the way that they were used since 30000BC. Moreover, the qualities of biometric features and the different kinds of biometric features and systems are analyzed. A few words are also said about the errors that can be encountered while using systems like that.

Chapter 3 Keystroke Sensitive Password Authentication Mechanism

The password authentication system that has been developed by using keystroke dynamics is presented. Firstly, the statement of the problem is reviewed, by issuing the biggest challenges that were faced in order to complete and implement the project to work correctly. The methodology and the classification algorithms that were used are also explained thoroughly.

Chapter 4 Mechanism Implementation

The chapter begins with a brief review of the programming environment that was used to create the authentication mechanism. In continuation, the implementation and the classes that were used for the training data collection and the classification are carefully explained, along with screen-shots of the interface and the run-time environment of the mechanism that was developed.

Chapter 5 Mechanism Evaluation

The testing environment is presented, from the number of subjects that were used to the number of repetitions that each subject typed, and the accuracy results. To conclude the

experiment, various findings are analyzed in order to better understand in what way each phase of the mechanism affected the results.

2 Biometric Systems

In this chapter, the term and the importance of identity management is introduced. Moreover, the different classifications of biometric systems are presented along with the phases in which they work, the different categories of features that can be extracted and used to authenticate or identify a certain individual, and the types of errors that might be encountered while using a biometric system.

2.1 Identity Management

Identity management encapsulates everything regarding the life cycle of a person's or an object's identity. This term includes many different aspects regarding identity, such as authentication, but also issues concerning trustworthiness, security, availability and portability. In more technical terms, identity management relates to the information, from a variety of procedures, which are required for specifying an entity as unique, in such a way to make it distinguishable among others.

In the commercial and technological world, the term identity management includes certain processes, methods and technologies that simplify and control the accessibility of applications and information, so it can protect them from unauthorized users. More specifically, in the commercial world, identity management suggests the combination of technological and operational procedures, in order to:

- Achieve a higher level of electronic entrepreneurship.
- Simplify the integration of enterprise applications.
- Control the flow of users that sign in, use and log out of the system.
- Keep statistical data about the use of certain services.

2.2 Identification Characteristics

The identification of a person can be done based on a variety of their characteristics, which are distinguished into:

- *Something the person knows*, including personal information, such as a password to log into the system
- *Something the person has*, including a distinguishing object, such as a card or an online certificate etc.
- *Something that the person is or does*, referring to information that is bound directly with the person, such as biometric features

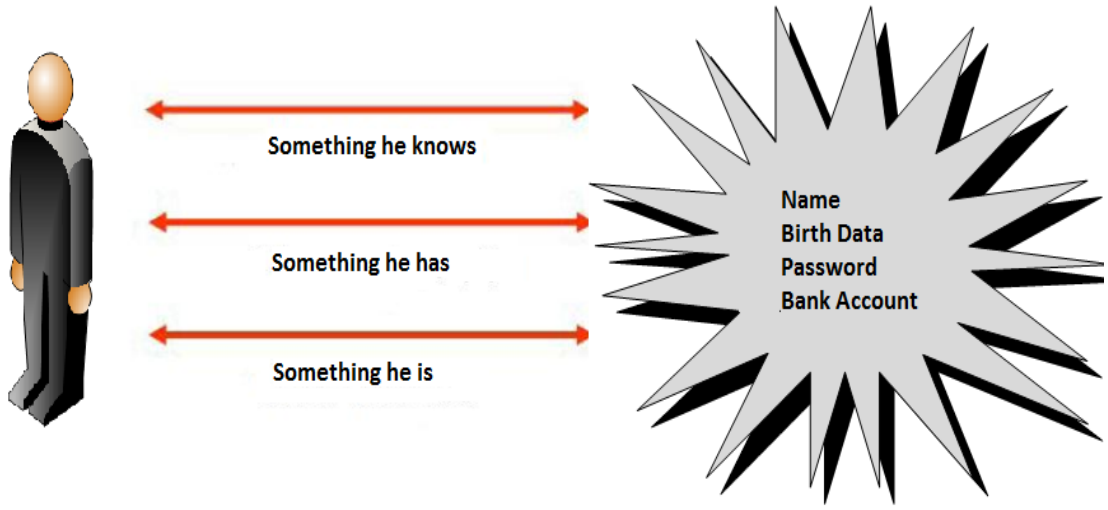


Figure 1 Means for Identification

The difference between those characteristics lies in their practicality, their application costs, their trustworthiness, their efficiency, the security they offer, and their level of respect for the person's privacy.

2.3 Historic Review of Biometrics-based Identification and Verification

The use of biometric systems to identify and distinguish people from each other is not something new. Since the beginning of human civilizations, people relied on the human face in order to

identify the people they knew and the ones they did not. In time, the identification procedure became harder, due to the population increase and the development of new traveling methods, which brought many unfamiliar faces to small societies.

In the course of time, a variety of human characteristics were used as more formal means for identification. For example, drawings were discovered in a 31000 year old cave, and it is said that they were made by the prehistoric people that lived there. Around the drawings signs of human hand prints were also found, which might have acted as signatures of the drawing's creators [10]. Signs also demonstrate, that fingerprints were used as an identification means around 500 BC when the Babylonians used to print their transactions on boards made out of clay, along with their fingerprints [11]. The Spanish explorer and writer Joao de Barros, wrote that in China of the 14th century, merchants used fingerprints the same way, in order to transact with each other. They also relied on footprints, which they printed on ink, in order to distinguish the children [12]. According to the same author, ancient Egyptian merchants relied on physical descriptors for the distinguishing of trusted merchants, with known reputation and previous successful transactions, from new traders.

In later years, the use of biometric features was observed more intensely. In 1890, Alphonse Bertillon, a French policeman, studied body mechanics and measurements in order to capture criminals [13]. The police relied on that method, known as the Bertillonage method, until the false identification of some people. The method was quickly abandoned and was reintroduced by Richard Edward Henry of the Scotland Yard.

In the beginning of the 20th century, Karl Pearson, an applied mathematics researcher, studied biometry in the University College of London. During his research, he made important discoveries in his field of study, through the study of statistical history and correlation, which he applied in animal evolution. His work included the method of moments, the Pearson distribution system, correlation and the chi-squared test.

Later, during the 1960's and 70's, biometric authentication procedures through signature were developed even further, but the biometric field remained without any important developments, until the military and security offices decided to continue the research on biometric technologies beyond just fingerprints.

In recent years, with the ever growing effort to improve identification and authentication

systems - which are mainly based on passwords and cards - along with a series of problems that were presented in the previous chapter, intense activity is shown in the field of biometric systems. An important application of these was the Super Bowl in Tampa, Florida, in 2001 [14] where a face recognition system was used on the faces of 100000 viewers, that were recorded by security cameras, in order to match them against the electronic images of criminals that the Tampa police provided. That move, combined with the unsuccessful operation of the system, caused many reactions from social freedom supporters, who accused biometric technologies as enemies of the personal rights and freedom.

Despite all the negative reactions, and, especially, after the attack on the 11th September of 2001, authorities all over the world continued with the installation of biometric systems mostly in airports, in order to identify "terrorist" suspects. A few years later, and after the continuous bomb attacks on underground railways, Great Britain adopted face recognition technologies in London, which, as of today, has over 200000 cameras in public areas all over the place for security reasons.

2.4 Qualitative Characteristics of Biometrics

As it has already been mentioned above, biometric technologies for identification, identity authentication and individuals' profiling are based on certain human characteristics. At this point a question arises about which human characteristics can be used to achieve such goals. In fact, any physical or behavioral human characteristic can be used, as long as it meets the following requirements [1]:

- *Universality:* every person needs to have this characteristic;
- *Distinctiveness:* the characteristic has got to be sufficiently diverse on every two people;
- *Permanence:* the characteristic has got to be sufficiently invariant (in relevance with the matching criterion), over a period of time;
- *Collect-ability:* there need to be methods allowing for the quantitative measurement of the characteristic.

Apart from the above, while taking in account the practical issues in the efforts of developing a biometric system, there are other issues as well, that should be taken under consideration, such as:

- *Performance:* which relates to the possible accuracy and speed of the recognition that we can achieve, and to the operational and environmental factors that influence those two
- *Acceptability:* which shows the extend to which people are willing to accept the use of a particular biometric identifier in their daily lives
- Circumvention: which reflects the level at which the system can be fooled by a malevolent user

2.5 Biometric Features and Systems Categorization

The huge variety of biometric features has led to their categorization. The two main categories on which biometric features are divided are *physiological (or passive)* and *behavioral (or active)*. The physiological biometric features include human characteristics that are generally stable and commonly unchanged such as fingerprints, hand geometry, the iris, the voice and the face. The behavioral biometric features refer to skills or functions performed by a person at a given time and for a specific reason, such as signing a paper or typing text on the computer.

Another classification of biometric features distinguishes them between *static* and *dynamic*. The term “static” refers to measuring an attribute that does not require any particular action at the time of the identification or the authentication. With the term “dynamic” we refer to the measurements that are recorded during a certain activity. For example, a signature can be measured either statically – by analyzing the signature after its collection – or dynamically – through the observation of the signing process.

Depending on the degree of separation between two individuals that they allow, but also depending on their permanence, biometric traits can be divided into *soft* and *hard* biometrics. Soft biometrics regard characteristics that provide some form of information about a person, but not sufficient for the effective separation between two people. These features can be continuous (such as height or weight) or discrete (such as sex, eye color, ethnicity). Hard biometrics contain all the information required to distinguish two people from each other. A person's DNA

[22][23][24], or their fingerprint are two basic examples of hard biometrics. The use of soft biometrics to identify or confirm a person's identity is only possible through their combination with hard biometrics. Important, however, is their usefulness on other applications, such as the connection of a person to a profile, while maintaining anonymity as we shall see in the next section.

The main categorization of biometric systems, which is found in the bibliography, is the separation of unimodal and multimodal systems. The first category includes biometric systems which are based on a biometric feature, such as, for example, facial recognition systems. Multimodal biometric systems combine more than one biometric feature to improve the accuracy of the system in terms of identification, authentication, or profiling. Examples of such systems are ones that combine face with voice recognition. In recent years, the term intramodal has also being used, which refers to biometric systems that focus on a biometric characteristic and include a fusion of different techniques at various stages of the biometric system.

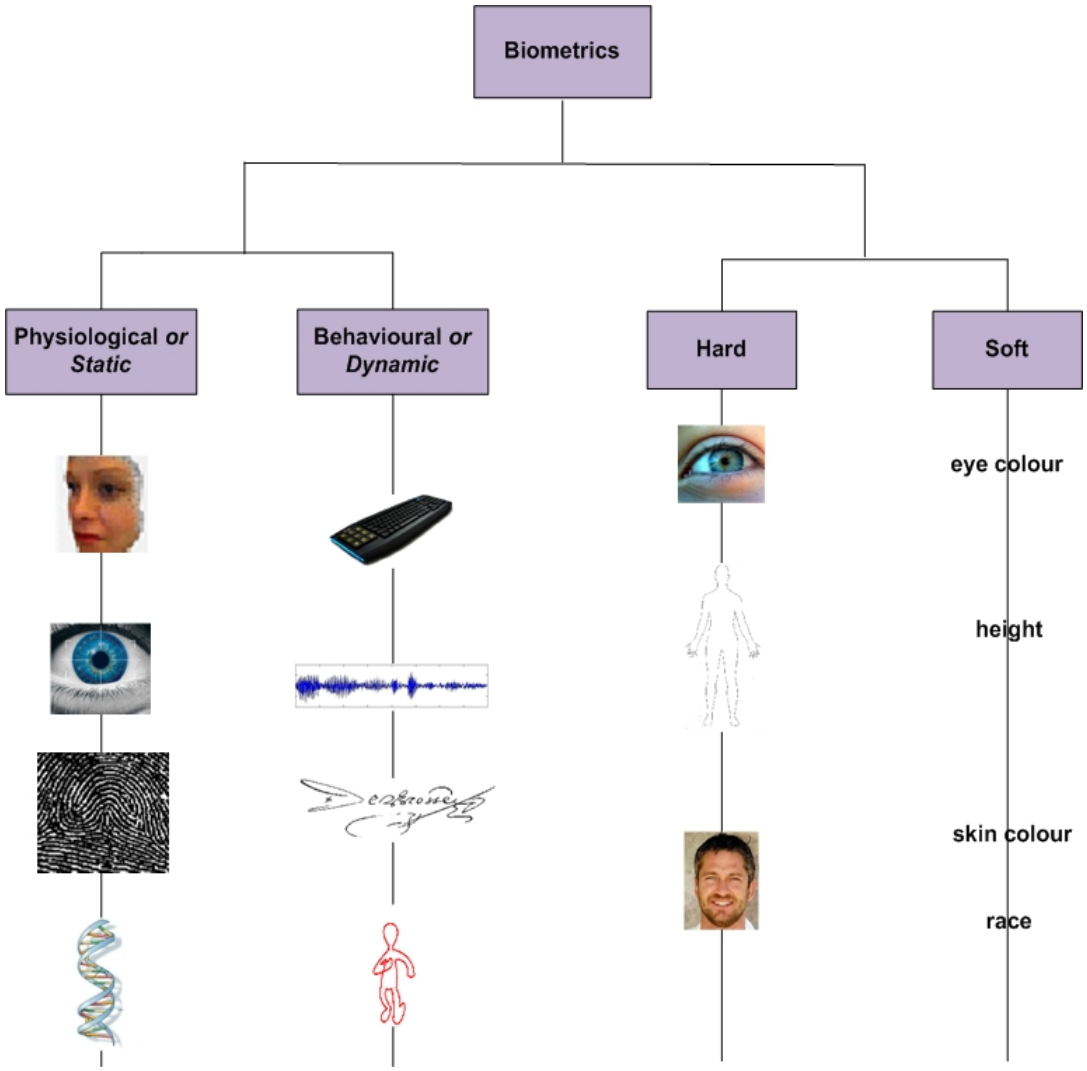


Figure 2 Classifications of Biometric Features and Systems

2.6 Biometric Features: Examples

2.6.1 The Face

Face recognition is a sub-field of the general problem of object recognition, which requires the separation between objects that differ little from one another, and for this reason, it comprises a major challenge in the field of computer vision. It is a non-invasive biometric method that can be applied in various areas such as trade (entertainment, video processing, videophones, video-conference), security, industries and law enforcement.

The first and most important step of a face recognition system is the detection of the face in the

picture. One image may contain none, one, or more faces. Thus, the face detection process aims at the automatic and reliable determination of the location of all the areas in the image which a person’s face, regardless of their three dimensional positioning, orientation, location and the light conditions. The complexity and efficiency of face recognition depend on the accuracy of the detection process on the image, which makes this image processing step essential. Any method that is used to detect a face on an image must take into account the large variability in terms of size, color, shape, texture, orientation, level of occlusion and expression of the human face.

The detection of a person's face in an image is a problem for which there has been hectic research activity, resulting in a huge variety of detection methods, which can even be categorized into four basic categories [26]:

- *Feature Invariant methods*: which aim at detecting the facial features that exist even when the angle, orientation, or lighting conditions vary. The detected characteristics are used to determine the position of the face. These methods are based in the observation that people have the ability to detect faces under different conditions and orientations, so there must be characteristics or properties that are not affected by these changing conditions and situations. Examples of such characteristics are the texture of the face (including hair and skin) and the color of the face. The last one is mainly used because studies have shown that the color difference of the face is manly due to the intensity rather than the chrominance, which results in a variety of color models of the human face which have been developed in various color spaces. In recent years, many techniques that combine facial features to identify faces have been developed. These features include the color of the skin, shape and size to determine the areas of the image that we need, followed by a confirmation phase on those areas of the picture, in which the existence of facial features is examined, such as nose, mouth, eyebrows and hair.
- *Knowledge-based methods:* which are based on the effort to decode the human knowledge, concerning the way that an ordinary human face looks like. Most often, this knowledge consists of rules that reflect the relationships between facial features and their relative distances and positions. An example of these rules is that a person has two eyes that are symmetrical to each other, a nose and a mouth. However, there is a risk. If those rules are too strict they might not be able to detect a face that does not completely satisfy

them, or if they are too general, they might provide false positives. In addition, detecting faces with different positions, increases greatly the number of cases that we have to deal with. However, in the case of images containing the front of the face or faces on a fixed posture, these methods can be quite efficient at detecting human faces.

- *Appearance-based methods:* which are based on models that are derived from a large set of face pictures with varied facial features, which are used for training, and result into a set of facial features that are stored either in the form of distribution models or in the form of a discrimination function. These models are used to detect faces by applying them on certain images.
- *Comparison-based methods:* which include the comparison of the input image, with standard face patterns that describe either the entire face or facial features separately. During this process, a calculation of correlation values takes place, and thus, the presence of a face in the input image is determined by those values.

Face recognition takes into account the result of face detection, and aims to identify or confirm the identity of the person detected on the image. The existing face recognition methods can be divided into two main categories: analytical and holistic methods [25] [28]. The analytical or feature-based methods focus on the study of feature extraction and calculate a set of geometrical features of the face such as eyes, nose and mouth. The holistic or appearance-based methods take into account the properties of a human face altogether. The face is recognized as a whole, without the use of certain points that could help distinguish the different areas of the face. These methods generally work directly on a pixel intensity presentation table, without the detection of facial features. Since no geometrical detection of the characteristics is needed, most of those methods are usually more practical and easier to implement than geometrical feature-based methods. Research efforts to combine analytical and holistic methods have also taken place [27].

2.6.2 Fingerprints

As already mentioned, the idea that fingerprints can be used to achieve identification and authentication of a person's identity was introduced centuries ago. The scientific breakthrough that occurred in the 20th century provided fertile ground for the development of automatic fingerprint identification systems which were first introduced and tested in the USA and

Australia. This first introduction of such biometric systems opened the way for broader use of fingerprints as the means of identification and authentication that are found in numerous places, such as, for example, official identification of people by police authorities or access control to computers and other gadgets.

The fingerprints are not genetically determined, but start growing randomly during the 3rd month of pregnancy on the fingertips of embryos [16]. The fingerprint identification techniques are based on these various structures at the end of human fingers. The uniqueness of fingerprints is based on the distinctiveness of various other characteristics, such as the pores, dots and incipient ridges. Thus, for example, the number of pores, their sizes, shapes and positions are factors that enhance the uniqueness of these features.

Taking a closer look at the morphology of the fingertips, pores are created where sweat glands in the subcutaneous layer of the skin generate sweat ducts. The latter ones grow through the subcutaneous layer and dermis to the epidermis, where the open duct on the skin’s surface presents itself as a pore [19]. Many factors influence the formation of the skin pattern [20], as is the stabilization that occurs when the secretion pores of the sweat glands open to the surface, at regular intervals. These openings on the surface are pores and their creation rate is an important parameter that determines the uniqueness of the composition of those pores. The position of the pores is stabilized once their formation is complete. Studies have concluded that pores do not disappear, move or are created over time [21].

Depending on the system, there are seven to nine typical patterns [17], which can be generally categorized as arch, whorl and loop, as shown in Figure 3. In addition to the main features of fingerprints, there are also secondary features such as the position of ridge ends and ridge bifurcations.

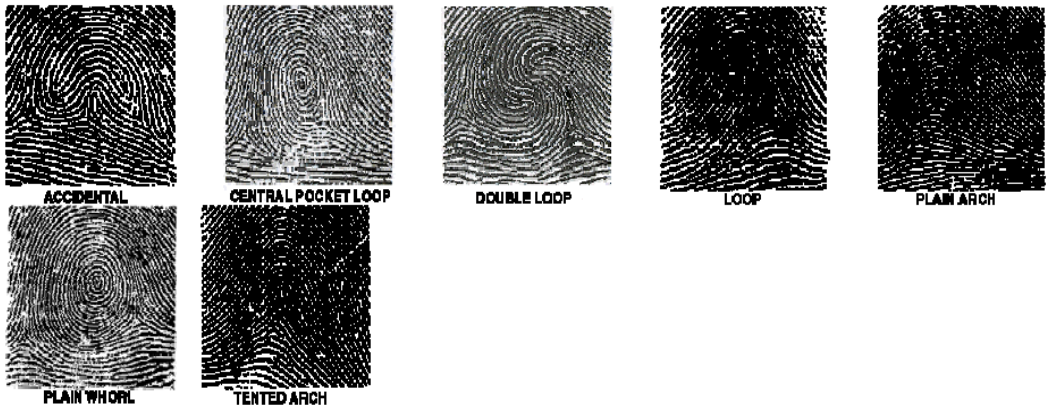


Figure 3 Seven common patterns used by the FBI [18]

Fingerprints can be taken with a variety of sensors, such as optical, electromagnetic, thermal and ultrasonic sensors. As we will also see in recent works [49] [50], new advanced sensors can detect if the fingerprints data is taken from a living organism (liveness detection) and have been developed to avoid cases where fingerprint identification systems have been "cheated" with the presentation of false or stolen fingerprint.

2.6.3 Keystroke Dynamics

The past two decades, computers have been a very important part of our lives and our society. The more we use them, the harder it gets to neglect the need of trustworthiness and further security on the applications and services that we use. Each day, we use computers to access our emails, browse forums, play video games, but also to manage our bank accounts and other sensitive data. Thus, we understand that a compromised account might cause a lot of trouble, and not only to the account's owner. That is where the keystroke dynamics biometric comes in.

In recent years, researches have shown that every person has certain unique features which are calculated by his keyboard typing rhythm [31]. The delay between each keystroke or the duration that he holds a certain key pressed while typing can distinguish him between other users. It is a behavioral biometric that can be used in many ways such as to detect emotions, but most importantly, to authenticate or even identify a person.

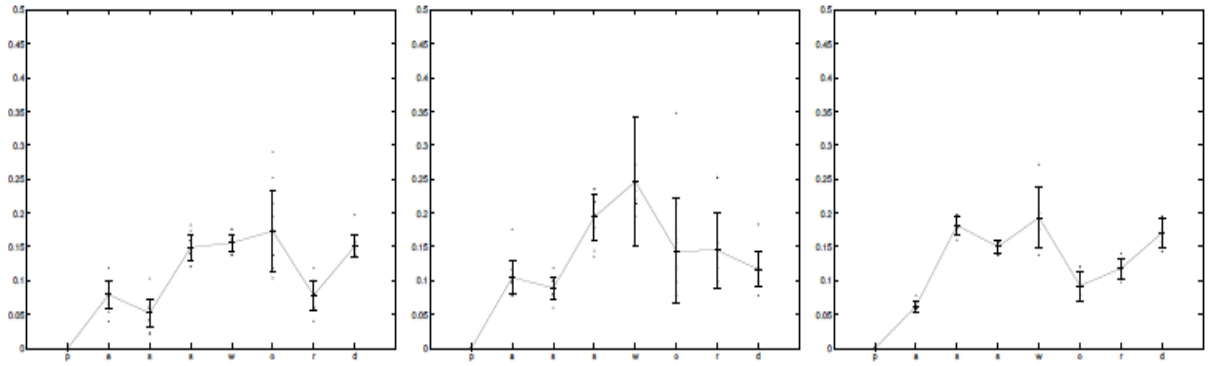


Figure 4 Latencies between keystrokes when writing the word “password” by three different people. [48]

For example, it can be used in a password authentication procedure. The password system is one of the most widely used ways to authenticate a person, especially on the internet. That is why keystroke sensitive password authentications can increase the security of a system dramatically. Imagine an authentication system that does not only require the correct password, but also the correct typing rhythm to accept the user. It acts the same way as the signature authentication in handwriting biometric features.

2.7 Biometrics: Applications

Biometric applications can be generally categorized as:

i. **Governmental control authentication models**: for these applications a public authority undertakes the task to collect biometric data in order to include them in authentication applications, such as identities, passports and social security cards. The control over such data can be centralized or it can be dispersed across several organizations, under the necessary agreements related to the disclosure or transfer of biometric data.

ii. **Access control model**: In this category a public or private authority is responsible for the collection of biometric data in order to ensure the access to a website or a web application. The control over the data can be either centralized or assigned across several organizations, with the appropriate agreements as before, or it can be distributed so that the person can share control.

iii. **Mixed model**: in this model, the collected biometric data are shared and exchanged between public and private authorities.

iv. **Facilitation model:** either the subject takes entirely the decision to use biometric data for personal facilitation use, or an organization uses the biometric data to simplify the management process either with a central or shared control.

v. **Monitoring model**: a public or private authority undertakes the task to collect and process biometric data for tracking purposes.

2.8 The two Operational Phases of Biometric Systems

All biometric systems operate in two distinct phases; capturing and storing the biometric reference data (enrollment) and querying the biometric system. The first phase comprises the initialization step of the biometric system. During this phase the system is provided with samples of the biometric characteristics of each individual in order to build a biometric reference template. For physiological biometric features, the system only needs to scan the feature and tie it to the certain individual. In the case of behavioral biometric features, the procedure is a little more complicated, because the system needs to train itself from a set of data. For example, in the case of keystroke dynamics, the individual needs to input his password a number of times, depending on the system. The system will then calculate certain mean vectors, which will later on be used in the classification phase.

In the second phase, the system compares the original data set with the one that is provided to access the system. Different algorithms are being used, depending on the biometric features, to determine if the given features satisfy the conditions and will be accepted by the system or not.

2.9 Types of Errors in Biometric Systems

System and statistical errors of the calculated data and the algorithms that are used extract the biometric features and compare the data, define the quality of the application of a biometric system, and it's ability to distinguish two different users from each other. A set of metric have been defined, in order to calculate the accuracy and the efficiency of biometric system. Those metrics represent the types of errors that might take place in the whole procedure of a biometric system.

- **False Acceptance Rate (FAR)**

  The false acceptance rate represents the probability that an authentication system might provide access to an impostor. The probability of that happening in an authentication system that works with physiological biometrics can be very low, but in the case of behavioral biometrics, it is not something unusual, depending on the strictness of the system. For this classification of biometrics in particular, this risk is intensified if the impostor has had the chance to watch and learn how to imitate the original user's way of authentication.

- **False Rejection Rate (FRR)**

  The false rejection rate represents the probability of an authentication system denying access to the original user. Again, such a risk is expected to be higher in the case of a behavioral biometrics based system than with physiological biometrics; especially if the user is not in his/her normal mental and emotional state, such as him being tired, in emotional distress, or under the influence of alcohol etc. It also depends on the strictness of the system.

- **False Match Rate (FMR)**

  The false match rate represents the probability of an identification system having a false positive match for a certain individual. It mostly depends on the algorithms that are being used, and the complexity of the biometric features that are being used. For example, face recognition and identification is one of the most complicated procedures, especially when the positioning, orientation, location and light conditions are not controlled.

- **False Non-Match Rate (FNMR)**

  The false non-match rate represents the probability of an identification system not being able to correctly identify the individual, while having his/her biometric data. Just like the false match rate, it strongly depends on the algorithms used, the complexity of the biometric features and the external conditions when data was captured.

- **Equal Error Rate (EER)**

The strictness of the system can be changed by the statistical error rates and algorithms that are used, depending on what we want to achieve, and what we deem important in an authentication or identification procedure. In the authentication procedure, we might, for example, not care about the false rejection rate, because we want our system to be extremely safe. Or we might not be very interested in keeping the system too safe, and decrease the strictness in which the system accepts a user. Same goes for the identification procedure, the FM and the FNM rates. The probability at which both types of errors FAR, FRR, and FMR, FNMR accordingly, have equal rates, is called the equal error rate; when for example both the FAR and the FRR are 3%.

One last important thing that we should know is that FAR, FRR and FMR, FNMR accordingly have inverse relationship. In an authentication system for example, the securer it is, hence the lower the FAR is, the higher the FRR will be as well, and vice versa.

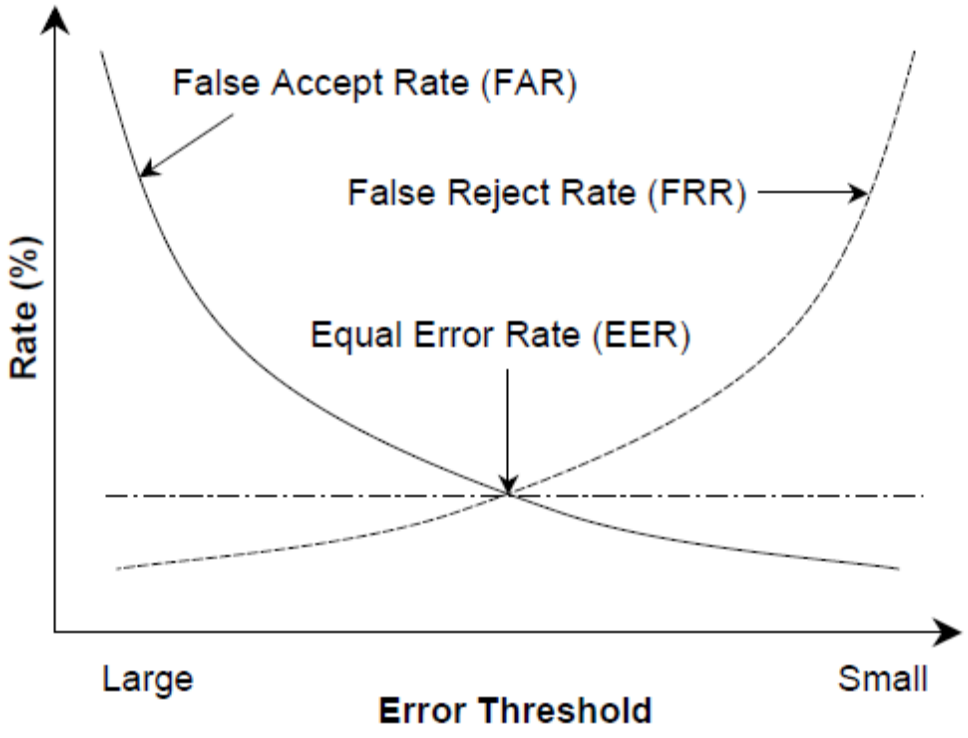


Figure 5 Relationship between FAR, FRR, and EER. [47]

3 Keystroke Sensitive Password Authentication Mechanism

This chapter begins with the statement of problem of keystroke dynamics, by presenting the methods used by studies that have been done through the past 20 years and the accuracies that they achieved. The methodology that was used on the mechanism is also explained, and the way the training and classification theoretically work.

3.1 Statement of Problem

Heavy research has been held during the last 30 years on keystroke dynamics, resulting among others in different classification algorithms and methodologies. The most common methods are based on traditional statistics, by using mean typing vectors and standard deviations [30] [32] [33] [34]. Over the years, many pattern-recognition methods have been applied to keystroke dynamics, such as neural networks [35] [36] [37], fuzzy logic [38] [39] [40] [41] [42], and support-vector machines [43] [44] [45] [46], among others.

In every study, a classification algorithm is proposed, a data set is collected, and then the classifier is used, so we can evaluate its accuracy. The accuracy results are often expressed in errors rates, and to be exact, in false acceptance (FAR) and false rejection rates (FRR). As we mentioned before, the FAR represents the chance that an authentication system might provide access to an impostor; FRR represents the chance of an authentication system denying access to the original user. The equal error rate is also used (EER), which represents the percentage at which both FAR and FRR are equal.

	Source Study	Classifier	Feature Sets				Password	
			Return key	Down-down	Up-down	Hold	Length	Reps
1	Joyce and Gupta (1990)	Manhattan	✓	✓			N/A	8
2	Bleha et al. (1990)	Euclidean		✓			11–17	30
		Mahalanobis		✓			11–17	30
3	Cho et al. (2000)	Mahalanobis *k*-NN	✓		✓	✓	7	75–325
		Auto-Associative Neural Net	✓		✓	✓	7	75–325
4	Haider et al. (2000)	Outlier Count		✓			7	15
5	Yu and Cho (2003)	SVM	✓		✓	✓	6–10	75–325
6	Araújo et al. (2004)	Scaled Manhattan		✓	✓	✓	10+	10
7	Kang et al. (2007)	*k*-Means			✓	✓	7–10	10

	Source Study	Filtering		Testing		Results (%)		
		Users	Times	#Attempts	Updating	Threshold	Miss	False Alarm
1	Joyce and Gupta (1990)		✓	1		heuristic	0.25	16.36
2	Bleha et al. (1990)			1	✓	heuristic	2.8	8.1[a]
				1	✓	heuristic	2.8	8.1
3	Cho et al. (2000)	✓	✓	1		zero-miss	0.0	19.5
		✓	✓	1		zero-miss	0.0	1.0
4	Haider et al. (2000)			2		heuristic	19.	11.[b]
				2		heuristic	22.	20.
				2		heuristic	13.	2.
5	Yu and Cho (2003)	N/A	N/A	1		zero-miss	0.0	15.78
6	Araújo et al. (2004)			1	✓	heuristic	1.89	1.45
7	Kang et al. (2007)	N/A	N/A	1	✓	equal-error	3.8	3.8

Table 1 Different evaluation procedures used by studies. [29]

Although there has been massive research effort, keystroke dynamics are still not applicable. The accuracy results of the classifiers used are inconsistent, because of the complexity of the keystroke dynamic features. There are some factors that affect the classifier error rates, which cannot always be anticipated. While there have been a few attempts in understanding those factors, the findings are incomplete. For example, the test subjects might type some strong passwords, or just ordinary English words. The results might be really good in the one case, but not in the other.

Studies have also shown that there are different kinds of typists. When a typist, with the characteristics of, let’s say Group 1, tries to copy the typing rhythm of a typist that belongs in Group 2, the FAR might even drop to 0%. But if they both belong in the same group, it gets much more complicated. The groups might for example differentiate experienced typists from amateur typists, or typists that have higher hold durations in general.

One of the biggest problems in keystroke dynamics, and behavioral biometrics in general, is that the increase of the strictness of the system might for instance lower the FAR to 0%, but it

will also increase the FRR. So if we have a system with FAR 5% and FRR 10%, and we try to lower the FAR to 0%, the FRR might jump to 30% or more. It is up to the researcher to choose the level of “strictness” of the system, which will affect the error rates significantly.

3.2 Methodology

The experiment on the mechanism that was created consists of four steps. The first step is the training data collection phase, where every test subject types in a certain password a number of times, so the system can create a profile from his typing pattern. The subjects that were selected for this experiment have different levels of typing skills.

The password selection was tricky, because on one hand it is more realistic to let the users chose their own password, but on the other hand, data collection would be a lot more difficult since different impostors would be needed for each password. A relatively strong password was used for this experiment, which consists of letters, numbers, and punctuation. The password that was used is:

password98.!

It is considered strong, because it contains 12 characters, numbers, punctuation and the need to press the Shift button.

Each of the subjects must first get used to the password. Then the training procedure begins where the subject types the password 20 more times while the mechanism keeps the timing data. In the case the subject made a typing mistake and uses backspace, the whole password and timings are deleted and typed again from the beginning. The features are also calculated, which include the down-down(DD), down-up(DU), up-down(UD) and the up-up(UU) durations of all the keystrokes.

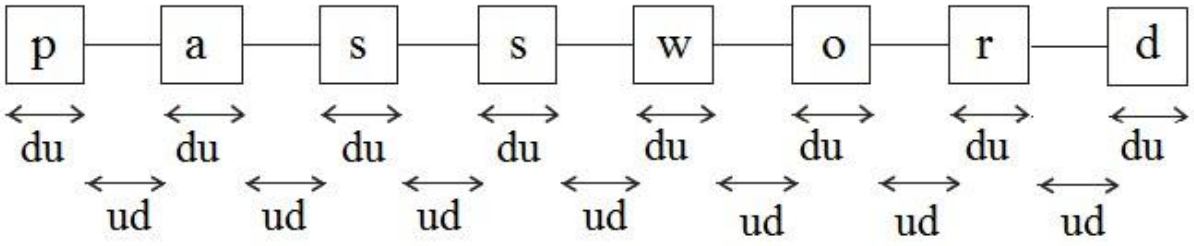


Figure 6 The down-up and up-down features of a password.

The second step in the evaluation was to implement the training, and classification algorithms that analyze password-timing data. The classification algorithms that were used are the Euclidean, the Manhattan and the Scaled Manhattan distance algorithms. They all simply create a threshold that will later affect the mechanisms decision to grant or deny access to a typist that tries to access the system.

In the third step, the FRR is calculated for each classifier, by having the test subject that trained the system, attempt to log in, while being watched by a second test subject who is going to act as the impostor and will try to assimilate his typing rhythm.

For the last step, the test subject that acts as the impostor, will try to access the system by trying to imitate the original users typing rhythm, so we can calculate the FAR for each classifier. Statistics are obviously kept throughout the last two phases of the experiment, so we can calculate the accuracy of the mechanism.

3.3 Classification Algorithms

The classifiers that were used are distance based and aim in distinguishing positive and negative attempts by representing the password entry samples in a vector space model. The classification begins with a training phase, in which the class’s centroid is computed in a p-dimensional space, where p is the number of features recorded for each sample. The distances between each training sample and the centroid are also calculated in order to define a threshold. The other phase of the classification is the testing phase, in which the actual

evaluation happens. If the distance between a test sample and the centroid created in the training phase surpasses the threshold, the mechanism will consider that attempt a failure.

The distances between the centroids and the samples are represented as *anomaly scores* in the following. The exact way the threshold is calculated is also explained later on.

- **Euclidean Algorithm.** The Euclidean distance classifier creates a p-dimensional space, where p is the number of features that are used, with each password being a point in that space. Hence, the training data is modelled as a cloud of points, with the mean vector acting as the centroid of the cloud. More specifically, in the training phase, the mean vector of each set of timing vectors is calculated and all the distances between the centroid and the training data points.

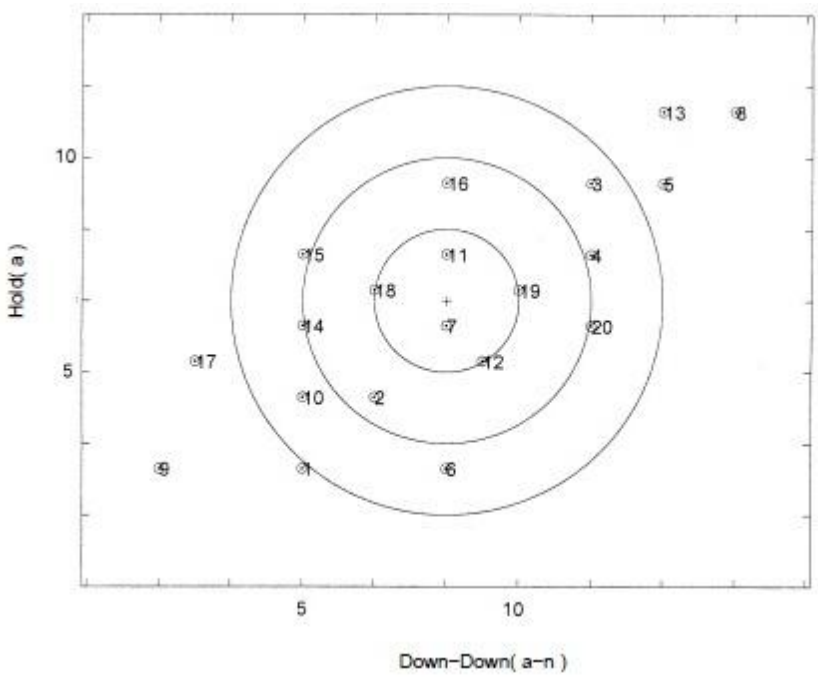


Figure 7 Contours of the decision surfaces for the Euclidean classifier.

In the classification phase, the Euclidean distances between the test vector and the mean vector are computed. If the distances are greater than the thresholds which were calculated in the training phase, the test vector is denied. The x_i value is representing the i-th timing feature of the mean vector and y_i is representing the i-th timing feature of the test vector, the anomaly score is calculated as

$$anomaly\ score = \sqrt{\sum_{i=1}^{p}(x_i - y_i)^2}$$

In Figure 7, we see the 2-dimensional space that was created by the hold (down-up) and the down-down features, and the contours which represent the way the Euclidean distance is measured. As noted above, the anomaly score of a point depends on its distance from that central mean vector. [29]

- **Manhattan.** The Manhattan distance classifier works almost the same way as the Euclidean classifier, except that the distance is calculated by using absolute values. In the training phase, the mean vectors of the timing vectors are calculated and the distances between the training data points and the centroid.

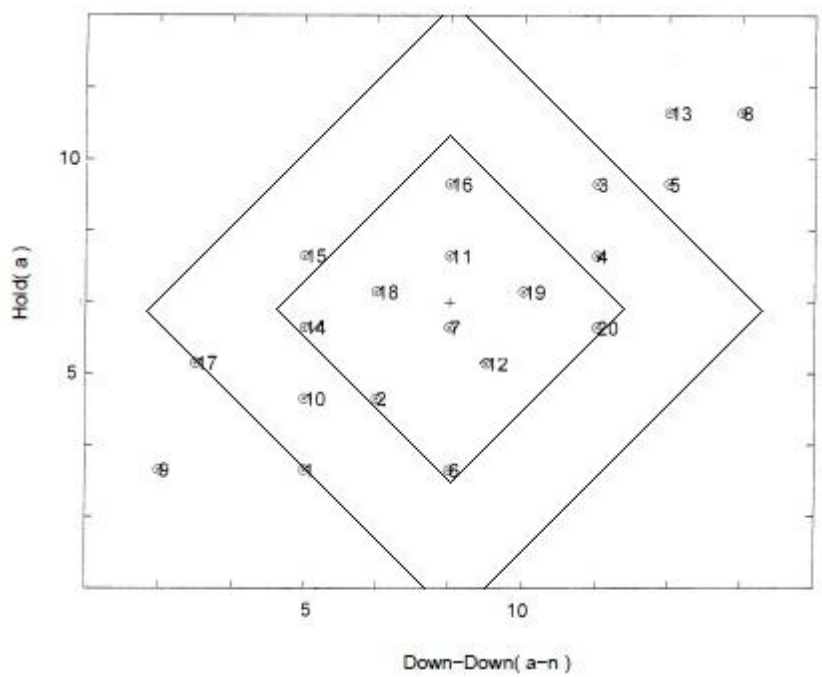


Figure 8 Contours of the decision surfaces for the Manhattan classifier.

In the classification phase, the Manhattan distance between the mean vector and the test vector is calculated. The difference between the two vectors is calculated across each of the features collected, and the absolute values of these differences are summed, same as

before. Again, the x_i value represents the i-th timing feature of the mean vector and y_i represents the i-th timing feature of the test vector, the anomaly score is calculated as

$$anomaly\ score = \sum_{i=1}^{p} |x_i - y_i|$$

In Figure 8, we the see 2-dimension space created by the hold and the down-down durations, such as before, but we see a difference in the contours of the decision surfaces because of the different distance formula. As with the Euclidean classifier, points in the center of the cloud have low anomaly scores, and points outside of the cloud, in the lower left, have high scores. Unlike the Euclidean classifier, points on the left edge of the cloud have relatively low anomaly scores. Because of the diamond shape of the Manhattan contours, points that differ from the center in only one dimension are assigned lower anomaly scores than those that differ in multiple dimensions. [29]

- **Scaled Manhattan.** The scaled Manhattan distance classifier is similar to the Manhattan classifier, but a scaling factor is also used, which is calculated by the set of training data. In the training phase, the mean vectors of the timing vectors are calculated, but also a vector of standard deviations which is going to work as the scaling factor.

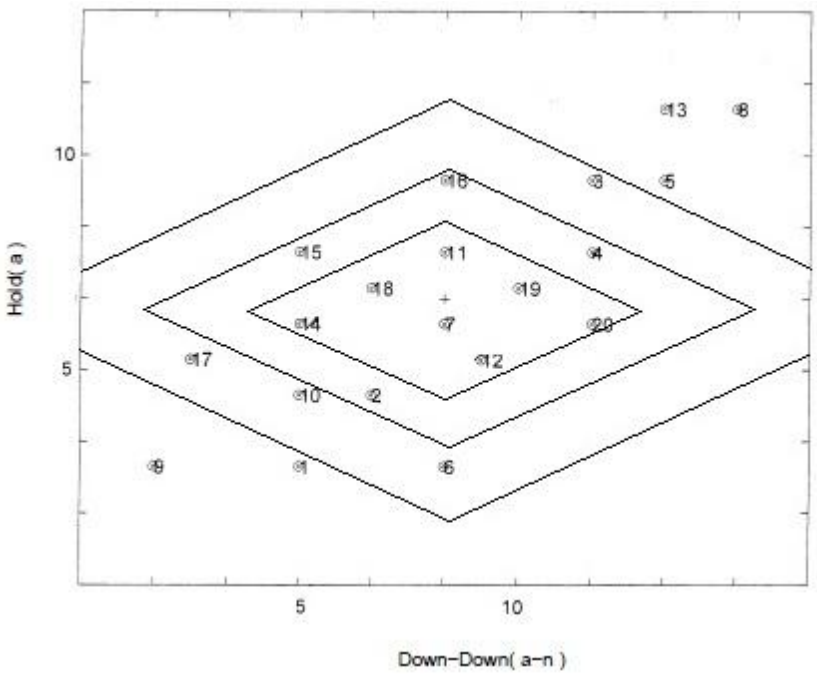


Figure 9 Contours of the decision surfaces for the Scaled Manhattan classifier.

In the classification phase, the calculation is similar to the Manhattan distance, but with a small difference. The distances for each feature are divided in each dimension by the standard deviation of the training samples that was calculated for that dimension. Specifically, with x_i and y_i defined as with the previous classifiers, and with s_i representing the standard deviation of the i-th feature of the training data set, the anomaly score is

$$anomaly\ score = \sum_{i=1}^{p} |x_i - y_i| / s_i$$

In Figure 9, we see the 2-dimensional space created by the features that we use, and the decision contours creating diamond shapes, a little wider than those of the Manhattan algorithm, before. The training data varies more widely in the Down-Down dimension than in the Hold dimension. Consequently, points on the right and left edge of the cloud of points have almost the same anomaly score as points on the upper and lower edge of the cloud. [29]

After the calculation of the anomaly scores is done, the mechanism can create the actual threshold. There are many different ways to create a threshold from a set of anomaly scores, depending on the strictness that we want to achieve. In this experiment, three different ways were used. The mechanism either:

- Calculates the mean vector of those anomaly scores, and uses that value as a threshold.
- The biggest anomaly score that was calculated is used with an error probability of 95%.
- In a sample of 100 anomaly scores for example, the 95th anomaly score is used, after sorting them from the lowest to the largest.

The first and the third case proved to be too strict in this mechanism’s case, so for the experiment, the second method was used.

4 Mechanism Implementation

4.1 Programming Environmental

The mechanism that is used for the keystroke dynamic authentication was programmed in the **JAVA** programming language.

The libraries that were imported and used are:

- **java.awt.event** Needed for the key-listeners, in order to record the timings between each keystroke.
- **javax.swing**
- **java.io**
- **java.util.ArrayList** Array-lists is the most common way to temporarily store data, when the size is not static, e.g., the password size.
- **org.w3c.dom** This library was used in order to parse the XML files containing the gathered data, and the anomaly scores.

In order to save the data, **XML** forms were used.

- **javax.xml.** Standard library used on JAVA, in order to parse XML files.

4.2 Training Data Collection

The training data collection procedure begins after the user has become familiar with the password and starts typing it in repetition through the Keystroke Trainer screen, as shown in

Error! Reference source not found.. The latter is in the “KeystrokeListener” class of the “keystrokelistener” package, extended by JFrame in order to create the *visual interface*, and has KeyListener and ActionListener implemented in order to *record* the keystroke actions. The KeyListener and the ActionListener are added on the password field as this is the field in which we need to keep track of the timings. More specifically, the events in which we are interested are the *press timings*, the *release timings*, and, in some cases, the *type of the key* that is being used. For example, when the “Shift” key is pressed, it is not being recorded. This choice is due to the fact that complications occurred in some cases, where test subjects might accidentally have pressed the shift key more times than they did on some of the repetitions that were recorded, which ended up in a logical error while trying to calculate the durations between each keystroke.

Saving the keystroke timings was quite complicated. That was mostly because some users might press multiple keys simultaneously before releasing any of them, which messed up the release times. In order to fix that, an array with “Button” items was used, which also saves the type of the key pressed or released. So, with a simple ‘if’, the mechanism can now determine which of the pressed keys is released, and add the release time to the exact item of the array.

The timings are all saved in milliseconds, in an array-list with “Button” items, which contain the two timings recorded and the order in which the keys were pressed. For example, the first key in our experiment, which is “p”, will be saved as “1”, the second, will be saved as “2” and so on. There is no need for the actual characters of the keys to implement keystroke dynamics. The password evaluation happens when the “OK” button is pressed as in every other program, the only data that is passed on to the next phase are the timings. Examples of correct and incorrect password evaluations are shown in Figure 11 and Figure 12. After a password repetition is entered, it is saved in a temporal array-list with “PasswordSample” items, which contain an array-list of “Button” items, so that we can create the raw data of the training sample, as shown in Figure 13.

When all password typing repetitions are over, the array-lists are saved in an XML form called “output” shown in Figure 18, which will be parsed by the “processDirectory” procedure of the “FeatureExtraction” class which is also in the “keystrokelistener” package of the program, in order to for the features to be calculated, which will be used for the actual training. The features

are calculated for each couple of keystrokes. For example, the first couple in this experiment is the keys “1” and “2”. The durations between their press times will be calculated (down-down), the durations between the press time and the release time of the first (down-up), the duration between the release time of the second and the press time of the first (up-down) and lastly, the duration between their release times (up-up). The features of one repetition are temporarily saved in an array-list with “Features” items, which contain four array-lists, one for each feature that will be used later on, in the classification phase. The “Features” are added in “FeatureSet” items, until the time that they are saved in XML format as well in the “featureOutput” form shown in Figure 19, in order to train the classification algorithms.

4.3 Classification

In this procedure, all the values needed for the classification are calculated, which obviously differ for the Euclidean, the Manhattan and the Scaled Manhattan algorithms.

The training process takes place in the “processData” class of the “keystrokelistener.training” package, which begins with “processDirectory” procedure by parsing the XML file “featureOutput” that contains the feature sets that were computed in the training data collection phase. All the contained data is temporarily inserted in “Features” and “FeatureSet” items that are found in the “keystrokelistener” package that was used for the training data collection. After the data is parsed, it is ready to be analyzed by the distance algorithms.

The Euclidean anomaly scores are calculated in the class “EuclideanTrainer”. The mean values of each feature are calculated, so that the anomaly scores can be calculated, by using the Euclidean distance formula explained in section 3.3. After the anomaly scores for each training sample are calculated, the threshold has got to be calculated. As explained in 3.3, there are three different ways to create a threshold, from which only one was used. Thus, to create a threshold, the *Collection.max* function is used, which returns the highest value from an array-list with numeric contents. That value is divided with 0.95 in order to achieve a 5% error probability according to classic statistics.

After a threshold is selected, it is added in a “EuclideanManhattanObject” item, which is used in both the Euclidean and the Manhattan trainer, because of their need for the same values for the testing phase. The values contained in those items are the anomaly score threshold and the mean vectors needed to calculate the anomaly score between the testing data and the centroid.

Same goes for the Manhattan anomaly scores, which are calculated in the “ManhattanTrainer” class. Firstly, The mean values are calculated and the anomaly scores between each training sample and the centroid with the help of the Manhattan distance formula shown in section 3.3 as well. A threshold is calculated from the anomaly scores that were calculated, just as with the “EuclideanTrainer”, and they are inserted in “EuclideanManhattanObject” items as well.

For the scaled Manhattan anomaly scores, which are calculated in the class “ScaledManhattanTrainer”, except from the mean values of the features, the standard deviations of each dimension have also got to be calculated, so that the scaled Manhattan anomaly scores can be computed by the Scaled Manhattan distance formula shown in section 3.3. After the anomaly scores are calculated, and a threshold is chosen, The calculated values are inserted in “ScaledManhattanObject” items, which contain the anomaly score threshold, the mean vectors and the standard deviations of each dimension.

After “EuclideanManhattanObject” and the “ScaledManhattanObject” items are returned to the “ProcessData” class, it creates an XML file for each anomaly detection algorithm, so that is three XML forms, the “euclideanStats.xml” shown in Figure 20, the “manhattanStats.xml” shown in Figure 21 and the “scaledManhattanStats.xml” shown in Figure 22, which will be used in the classification phase of the test samples created when someone attempts to evaluate his typing rhythm.

The last phase of the mechanism is the authentication phase, where the user will attempt to log into the trained mechanism. The “Authentication” screen shown in Figure 14 is created by the “LoginScreen” class, which is contained in the “keystrokelistener.authentication” package, it also extends the JFrame interface, and implements the KeyListener and the ActionListener, exactly as the “KeystrokeListener” in the training data collection phase. They also work alike, so when a user attempts to access the system and types in the password, his keystroke timings are recorded and saved in an array-list just as before, so that the features can be calculated.

If the password entry is correct, his features are calculated by the “calculateFeatures” class, and they are inserted in a “Features” item, which is contained in the “keystrokelistener” package that was used before, in the training data collection.

For the classification, all three classification algorithms are used, to create the anomaly scores between the test sample and the centroid, and if the scores are lower than the thresholds, the user will be granted access as shown in Figure 15, if not, he will be denied access as shown in Figure 16. More specifically, after the XML forms with the threshold data are brought up by the “getEuclideanManhattanProfile” and the “getScaledManhattanProfile” classes, the classes “euclideanAuthentication”, “manhattanAuthentication” and “scaledManhattanAuthentication” calculated the anomaly score between each feature of the training sample and the centroids, then compare that value with the anomaly score threshold, and return a ‘true’ or a ‘false’ value depending on the outcome of the comparison.

For each pair of keystrokes that the anomaly score is lower than the threshold, a counter is increased by one. Tests showed that the subjects usually show greater anomaly scores than the threshold on the first and the last pairs of keystrokes recorded. For this, if the counter is larger or equal to the size of the keystroke pairs minus two, the user will be accepted. If not, the user will be denied access to the system.

The use of three or more algorithms might prove helpful in the final conclusion of granting access to a user or not, despite the scientific interest of the research. In many cases for example, one detection algorithm might grant access to a user, while the other will not, as shown in Figure 17. So, the final decision can be made if the majority of the algorithms accept a user.

4.4 Run-time and Interface Screen-shots

In the following figures the interface built for users to interact with the keystroke dynamics biometric system is presented both in training and in authentication phases.

Figures 10 to 13 show the different case scenarios of the “Keystroke Trainer” screen, in the training data collection process. Details about the scenario’s situation are explained in each figure’s caption. In Figures 14 to 17, the different case scenarios of the “Authentication” screen, in the evaluation procedure are shown. Just as before, details about the scenario’s situation are

explain in each figure’s caption. Finally, in Figures 18 to 22 the different XML messages produced at different steps of the process are presented. Details about the form’s origin and purpose are on the figure’s caption.

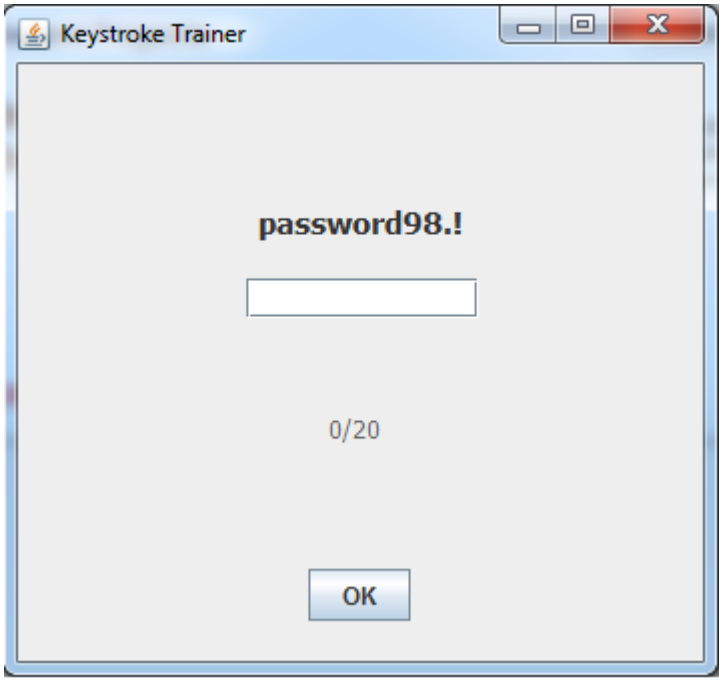


Figure 10 The Keystroke Trainer screen at Training Mode.

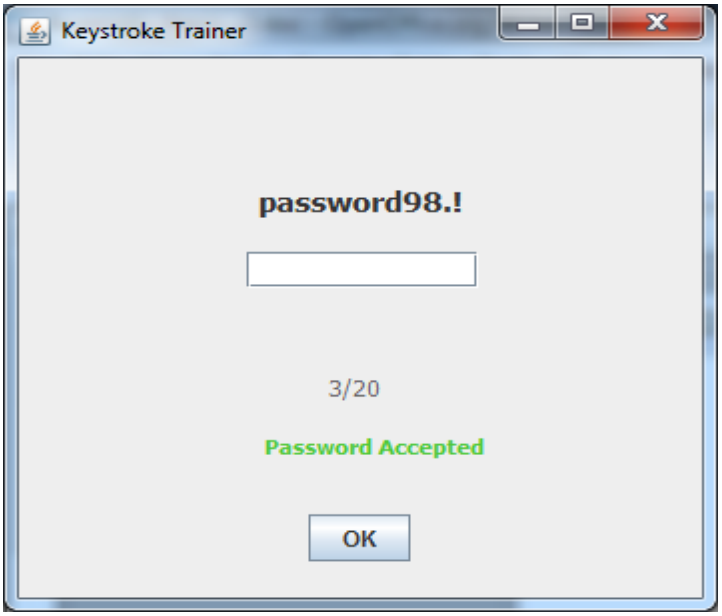


Figure 11 The “Keystroke Trainer” screen, after accepting a password sample.

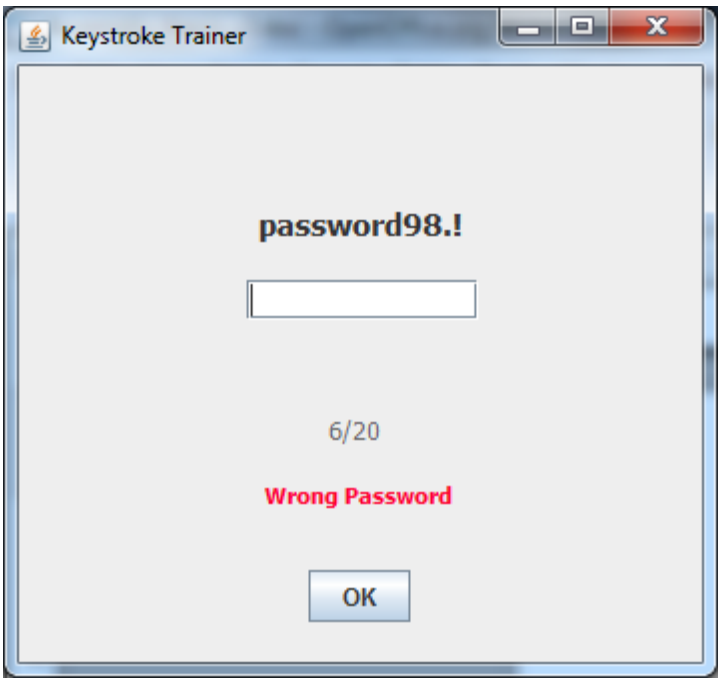


Figure 12 The “Keystroke Trainer” screen, after a wrong password entry.

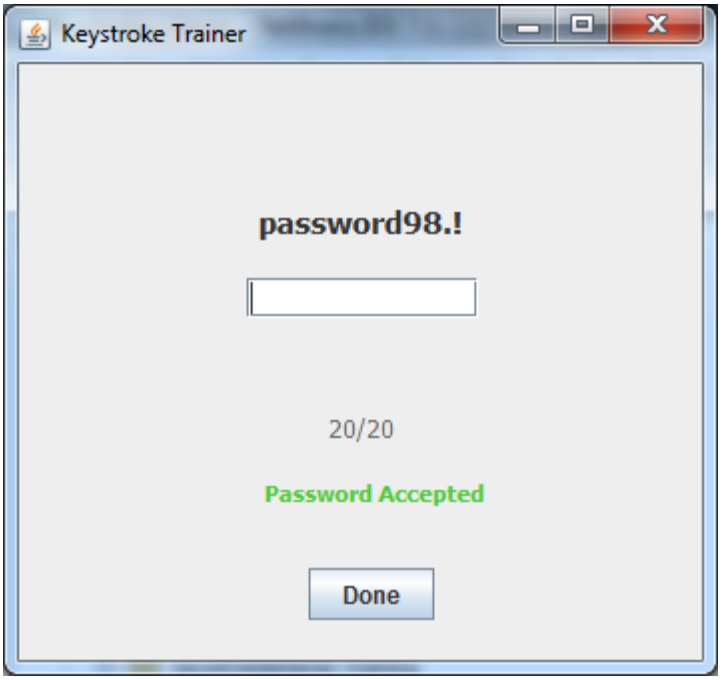


Figure 13 The “Keystroke Trainer” screen, upon completion of the training data collection.

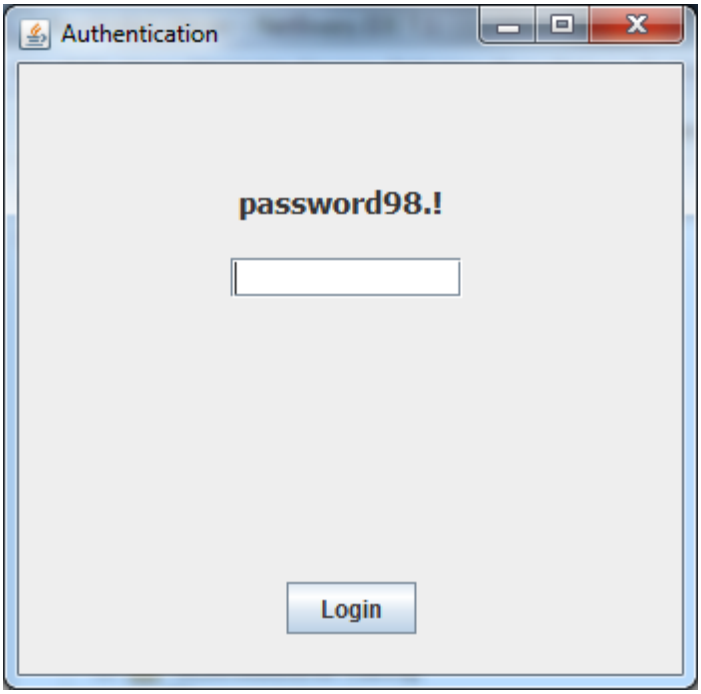


Figure 14 The “Authentication” screen.

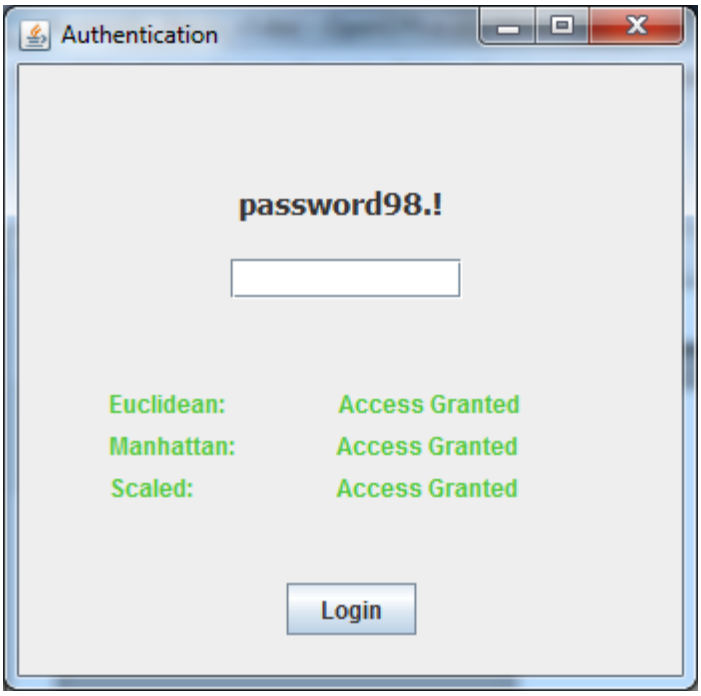


Figure 15 The “Authentication” screen, with all classification algorithms granting access.

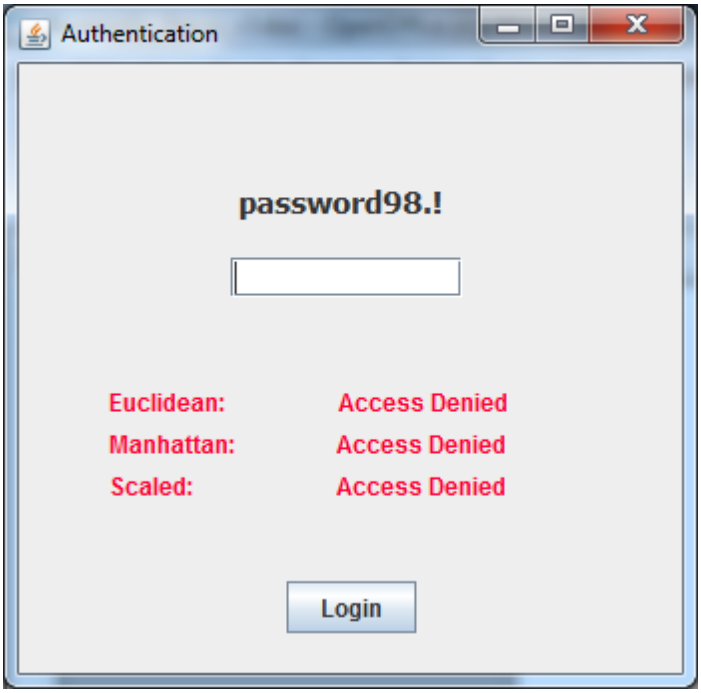


Figure 16 The “Authentication” screen, with all classification algorithms denying access.

Figure 17 The “Authentication” screen, with only one classification algorithm granting access.

```
- <user>
  - <password sample="1">
    - <button key="1">
        <presstime>1.347565492673E12</presstime>
        <releasetime>1.347565492823E12</releasetime>
      </button>
    - <button key="2">
        <presstime>1.347565492743E12</presstime>
        <releasetime>1.347565492923E12</releasetime>
      </button>
    - <button key="3">
        <presstime>1.347565492913E12</presstime>
        <releasetime>1.347565493033E12</releasetime>
      </button>
    - <button key="4">
        <presstime>1.347565493103E12</presstime>
        <releasetime>1.347565493253E12</releasetime>
      </button>
    - <button key="5">
        <presstime>1.347565493863E12</presstime>
        <releasetime>1.347565494033E12</releasetime>
      </button>
    - <button key="6">
        <presstime>1.347565494093E12</presstime>
        <releasetime>1.347565494263E12</releasetime>
      </button>
    - <button key="7">
        <presstime>1.347565494413E12</presstime>
        <releasetime>1.347565494573E12</releasetime>
      </button>
    - <button key="8">
        <presstime>1.347565494743E12</presstime>
        <releasetime>1.347565494903E12</releasetime>
      </button>
    - <button key="9">
        <presstime>1.347565495103E12</presstime>
        <releasetime>1.347565495223E12</releasetime>
      </button>
    - <button key="10">
        <presstime>1.347565495303E12</presstime>
        <releasetime>1.347565495403E12</releasetime>
      </button>
```


Figure 18 Sample of the “output” XML containing the raw keystroke data.

```
- <user>
  - <features set="1">
    - <button keys="1-2">
        <down-down>70.0</down-down>
        <down-up>150.0</down-up>
        <up-down>-80.0</up-down>
        <up-up>100.0</up-up>
      </button>
    - <button keys="2-3">
        <down-down>170.0</down-down>
        <down-up>180.0</down-up>
        <up-down>-10.0</up-down>
        <up-up>110.0</up-up>
      </button>
    - <button keys="3-4">
        <down-down>190.0</down-down>
        <down-up>120.0</down-up>
        <up-down>70.0</up-down>
        <up-up>220.0</up-up>
      </button>
    - <button keys="4-5">
        <down-down>760.0</down-down>
        <down-up>150.0</down-up>
        <up-down>610.0</up-down>
        <up-up>780.0</up-up>
      </button>
    - <button keys="5-6">
        <down-down>230.0</down-down>
        <down-up>170.0</down-up>
        <up-down>60.0</up-down>
        <up-up>230.0</up-up>
      </button>
    - <button keys="6-7">
        <down-down>320.0</down-down>
        <down-up>170.0</down-up>
        <up-down>150.0</up-down>
        <up-up>310.0</up-up>
      </button>
    - <button keys="7-8">
        <down-down>330.0</down-down>
        <down-up>160.0</down-up>
        <up-down>170.0</up-down>
        <up-up>330.0</up-up>
      </button>
```


Figure 19 Sample of the “featureOutput” XML, containing the features extracted from the raw data.

```
- <anomalies>
  - <euclidean>
    - <button keys="1-2">
        <threshold>76.33325372994393</threshold>
        <ddmean>66.0</ddmean>
        <dumean>120.5</dumean>
        <udmean>-54.5</udmean>
        <uumean>131.5</uumean>
      </button>
    - <button keys="2-3">
        <threshold>169.62222731705887</threshold>
        <ddmean>172.0</ddmean>
        <dumean>186.0</dumean>
        <udmean>-14.0</udmean>
        <uumean>127.5</uumean>
      </button>
    - <button keys="3-4">
        <threshold>220.07319521013912</threshold>
        <ddmean>208.5</ddmean>
        <dumean>141.5</dumean>
        <udmean>67.0</udmean>
        <uumean>235.5</uumean>
      </button>
    - <button keys="4-5">
        <threshold>393.38288679732875</threshold>
        <ddmean>406.0</ddmean>
        <dumean>168.5</dumean>
        <udmean>237.5</udmean>
        <uumean>401.5</uumean>
      </button>
    - <button keys="5-6">
        <threshold>226.49083866682113</threshold>
        <ddmean>262.0</ddmean>
        <dumean>164.0</dumean>
        <udmean>98.0</udmean>
        <uumean>224.5</uumean>
      </button>
    - <button keys="6-7">
        <threshold>79.86924079393768</threshold>
        <ddmean>266.0</ddmean>
        <dumean>126.5</dumean>
        <udmean>139.5</udmean>
        <uumean>277.0</uumean>
      </button>
```


Figure 20 Sample of the “euclideanStats.xml” file, containing the anomaly score mean vectors created by the Euclidean algorithm.

```
- <anomalies>
  - <manhattan>
    - <button keys="1-2">
        <threshold>107.82499999999999</threshold>
        <ddmean>66.0</ddmean>
        <dumean>120.5</dumean>
        <udmean>-54.5</udmean>
        <uumean>131.5</uumean>
      </button>
    - <button keys="2-3">
        <threshold>182.39999999999998</threshold>
        <ddmean>172.0</ddmean>
        <dumean>186.0</dumean>
        <udmean>-14.0</udmean>
        <uumean>127.5</uumean>
      </button>
    - <button keys="3-4">
        <threshold>228.0</threshold>
        <ddmean>208.5</ddmean>
        <dumean>141.5</dumean>
        <udmean>67.0</udmean>
        <uumean>235.5</uumean>
      </button>
    - <button keys="4-5">
        <threshold>401.375</threshold>
        <ddmean>406.0</ddmean>
        <dumean>168.5</dumean>
        <udmean>237.5</udmean>
        <uumean>401.5</uumean>
      </button>
    - <button keys="5-6">
        <threshold>239.39999999999998</threshold>
        <ddmean>262.0</ddmean>
        <dumean>164.0</dumean>
        <udmean>98.0</udmean>
        <uumean>224.5</uumean>
      </button>
    - <button keys="6-7">
        <threshold>92.625</threshold>
        <ddmean>266.0</ddmean>
        <dumean>126.5</dumean>
        <udmean>139.5</udmean>
        <uumean>277.0</uumean>
      </button>
    - <button keys="7-8">
        <threshold>136.325</threshold>
        <ddmean>239.0</ddmean>
        <dumean>137.5</dumean>
        <udmean>101.5</udmean>
        <uumean>267.5</uumean>
      </button>
```


Figure 21 Sample of the “manhattanStats.xml” file, containing the anomaly score mean vectors created by the Manhattan algorithm.

```
- <anomalies>
  - <scaledmanhattan>
    - <button keys="1-2">
        <threshold>4.578569706031895</threshold>
        <ddmean>66.0</ddmean>
        <dumean>120.5</dumean>
        <udmean>-54.5</udmean>
        <uumean>131.5</uumean>
        <dds>30.23243291566195</dds>
        <dus>19.615045245933032</dus>
        <uds>24.794152536434876</uds>
        <uus>24.14021540914662</uus>
      </button>
    - <button keys="2-3">
        <threshold>3.4230991758441576</threshold>
        <ddmean>172.0</ddmean>
        <dumean>186.0</dumean>
        <udmean>-14.0</udmean>
        <uumean>127.5</uumean>
        <dds>62.73754856543249</dds>
        <dus>18.275666882497067</dus>
        <uds>64.60650122085238</uds>
        <uus>73.81564874740315</uus>
      </button>
    - <button keys="3-4">
        <threshold>3.6828905370488205</threshold>
        <ddmean>208.5</ddmean>
        <dumean>141.5</dumean>
        <udmean>67.0</udmean>
        <uumean>235.5</uumean>
        <dds>64.82862022286145</dds>
        <dus>27.798381247835277</dus>
        <uds>55.95533933415113</uds>
        <uus>66.96827607158482</uus>
      </button>
    - <button keys="4-5">
        <threshold>2.885888435278091</threshold>
        <ddmean>406.0</ddmean>
        <dumean>168.5</dumean>
        <udmean>237.5</udmean>
        <uumean>401.5</uumean>
        <dds>157.27046766637403</dds>
        <dus>23.510635891017493</dus>
        <uds>159.30709337628377</uds>
        <uus>159.97734214569263</uus>
      </button>
```


Figure 22 Sample of the “scaledManhattanStats.xml**”** file, containing the anomaly score mean vectors and the standard deviations created by the Scaled Manhattan algorithm.

5 Mechanism Evaluation

In this chapter, the testing environment is presented, the number of test subjects that were used, the number of password repetitions to train the algorithms, and to evaluate the mechanism's accuracy. Also, screen-shots of the program's run-time environment and the XML forms are shown. The results of the experiment are presented, and lastly, a conclusion is made on ways to improve the mechanism in future work.

5.1 Testing Environment

In order gather the data and calculate the accuracy of the mechanism that was created for this experiment, 20 test subjects were picked randomly, with various typing skill levels. They were divided into groups of two, so when the one trained the mechanism with his typing pattern, the other acted as the impostor.

First, the test subjects had to get used to the password that was chosen for the experiment, by typing it multiple times. The number of times depended on each individual, and was usually about 15-20 times. Next, each subject entered the password 20 times in the “Keystroke Trainer” screen of the program, in order to train the three classifiers that were used in this experiment.

After the training process of the program, each of the users attempted to log into the system 10 times, while the subject who acted as the impostor tried and assimilated his typing pattern in an effort to copy his typing rhythm.

After the first user got 10 password repetitions done, no matter the FRR, the second user acted as the impostor by attempting to imitate the first user's typing pattern, 10 times as well so that the FAR could be calculated. And then backwards, the second user that acted as the impostor in the

first step trained the system, and the user that acted as the original one, would act as the impostor by trying to assimilate the other user's rhythm, and so on.

5.2 Results

	FRR Euclidean	**FRR** Manhattan	**FRR** Scaled Manhattan	**FAR** Euclidean	**FAR** Manhattan	**FAR** Scaled Manhattan
User 1	1/10	3/10	3/10	0/10	0/10	0/10
User 2	2/10	2/10	2/10	2/10	0/10	3/10
User 3	1/10	2/10	2/10	0/10	0/10	0/10
User 4	0/10	1/10	0/10	1/10	2/10	3/10
User 5	0/10	0/10	0/10	2/10	2/10	2/10
User 6	1/10	2/10	0/10	1/10	0/10	1/10
User 7	1/10	1/10	1/10	1/10	1/10	2/10
User 8	1/10	3/10	5/10	0/10	0/10	0/10
User 9	2/10	3/10	1/10	1/10	0/10	1/10
User 10	2/10	3/10	2/10	1/10	0/10	0/10
User 11	2/10	2/10	1/10	0/10	0/10	0/10
User 12	1/10	4/10	2/10	0/10	0/10	0/10
User 13	1/10	2/10	2/10	3/10	0/10	1/10
User 14	0/10	2/10	0/10	2/10	1/10	0/10
User 15	1/10	2/10	1/10	3/10	1/10	0/10
User 16	0/10	2/10	1/10	0/10	0/10	0/10
User 17	1/10	1/10	1/10	4/10	3/10	0/10
User 18	1/10	2/10	2/10	0/10	0/10	0/10
User 19	2/10	1/10	3/10	0/10	1/10	0/10
User 20	1/10	0/10	1/10	1/10	1/10	1/10

Total	10,5%	19%	15%	11%	6%	7%

Table 2 Accuracy Results of the Mechanism in terms of FRR and FAR

As we can see, the FAR and the FRR are so far determined by the strictness of each classification algorithm, which in our case is firstly the Manhattan algorithm, which is the strictest of them all, next goes the scaled-Manhattan algorithm, and last the Euclidean algorithm.

We also observe, that for some users, although the total FRR of the Manhattan distance algorithm, for example, is the highest, for some users it was the lowest, a phenomenon that shows us that there are unexplained factors that determine the strictness of each algorithm, which we will try to be explained in the next section of this chapter.

5.3 Conclusion

The conclusion of this experiment is that the implementation of keystroke dynamics will dramatically improve the security of password security systems, but not in the current state of the mechanism. The current mechanism was implemented in half a year's work, so, obviously, there is a lot of room for improvement in almost every phase of the mechanism.

Training Data Collection and Features. In the training data collection, and specifically in the feature extraction, there are many different choices about which features to calculate and use for the classification. For example, the mechanism calculates the down-down, down-up, up-down and up-up durations of each pair of keystrokes. But, results showed that the classification gets way too strict if all are used, as a p-dimensional distance. The FRR reached almost 100% for every single anomaly detection algorithms. So, in the end, only the down-down and the down-up features were used, modeling a 2-dimensional cloud of points. Thus, we understand that the number of features that are extracted from the raw timing data might change the outcome of the evaluation process.

Static or Dynamic Password. The fact that a static password was used for all test subjects influenced the outcome of the experiment as well. A normal experiment would require the test subjects to type hundreds of repetitions, in order to get more used to the password and to

establish a more stable typing pattern. But sadly, that was not an option for this experiment, because there wouldn't enough volunteers. If each test subject had the possibility to choose his own password, the results would obviously differ greatly. It would be a lot harder for the impostor to assimilate the original user's typing pattern, and the threshold would be a lot lower because of the more stable typing rhythm of the original user. But for the sake of the experiment, a static password was used, in order to demonstrate how each user's typing rhythm differs.

Classification Algorithms. The anomaly-detection algorithms that are used, are one of the most obvious factors in the result's outcome. As it has been shown by many researches, the FRR and the FAR are greatly determined by the algorithms used. For example, in this experiment, the Manhattan algorithm usually appeared to be the strictest of all others used. But that also depended on the user's typing rhythms, for some user's the strictest was the scaled-Manhattan, a phenomenon that cannot be explained in the current level of the research.

Threshold selection. As it was mentioned in 4.3, there are different ways to determine the threshold from the anomaly scores. In some cases the mechanism because stricter than it is ought to be, increasing the FRR too much. In some other cases, when choosing the greatest value of the anomaly scores calculated for example, the strictness of the evaluation process fell dramatically, by easily accepting the impostor, which ended up in a failed experiment attempt, because of a very high FAR.

Different Typing Rhythms. Another interesting finding of the experiment was the way the FAR increased with the typing rhythms of the test subjects. Someone would assume that it is easy to imitate a user that types slowly, but that was not the case. In fact, when an experienced user tried to imitate an amateur typist, the FAR got close to 0%. Seems that even amateur users, who might even type with two fingers, have very specific duration's between each keystroke, which are very hard to imitate. Same goes for users with very fast typing rhythms, but not only because the impostor was not always able to achieve their typing rhythm, but the research showed, that when the typist's rhythm is fast, his hold durations (down-down and down-up) get a lot more complicated. The couples with high FAR were usually the ones with average typing skills. The FAR appeared to also increase on certain couples, where the users had similar typing

rhythms, while someone with a different kind of typing rhythm could not achieve to imitate them and get access to the system.

In some cases, the FRR was high as well, but that was mostly because users did not get used to the password as good as they should have. One of the most important negative factors of using a static password, as was mentioned before. Some even forgot their typing rhythm, which made them unable to get access to the system after a certain point. Those cases are called false enrollments.

6 Summary

Within this research, the biometric systems were reviewed through the course of history and analyzed in many ways. Biometric authentication and identification has been around for a almost 30 thousand years, and continues to grow each day with feverish research taking place all over the world.

A biometric system is mostly based on the universality, the distinctiveness and the permanence of an individual’s characteristics. Factors like that are very important in order to distinguish two people from each other or to identify someone by a certain feature that he possesses.

Features can also be of various kinds. The most important division of features are of the physiological and of the behavioral categories. A physiological biometric feature can be anything regarding an individual’s appearance, like for example his face, or his jaw-line (teeth). Such characteristics are, in general, easier to be analyzed, although systems based on them inherit a series of scientific challenges. Behavioral features have got to do with a person’s behavior, obviously. For example with the way he does a certain act, for example the way he signs a paper with his signature. The common signature can tell people apart because of the different way each person hard-writes, which is developed through the course of his life, and is totally unique.

A very interesting behavioral characteristic of people are the keystroke dynamics, the rhythm that each person types on the keyboard, which is unique as well. Within this experiment, a mechanism was created using keystroke dynamic features in order evaluate the authenticity of a person while typing his password.

Much research has been made in the field of keystroke dynamics since the beginning of the use of computers, and keyboards more especially, so the last 30 years or so. Many different methods have been used in order to distinguish people from each other according to their typing rhythms, such as traditional statistics, neural networks, fuzzy logic or and machine-support vectors.

In this experiment, a mechanism was created, which used traditional statistics anomaly-detection algorithms, which included the Euclidean, the Manhattan and the scaled-Manhattan algorithms. The mechanism works in three different phases, *the training data collection phase*, *the training of the classification algorithms*, and *the classification itself*, in order to evaluate a user as the original one or as an impostor.

After the development of the mechanism, a testing phase took place, during which 20 individuals were chosen in order to determine the accuracy of the mechanism, which is measured in metrics such as the false acceptance rate (FAR) and the false rejection rate (FRR), which represent the probability of the mechanism denying access to the original user, and the probability of the mechanism granting access to an impostor, respectively.

After acquiring the results, a number of conclusions were made, about the way the experiment was done and about the mechanism itself. Things such as the way the training data collection was done, the features that were extracted and used, the classification algorithms, the threshold that was selected, the password that was used and the different kind of typing rhythms of the individuals, determined greatly the outcome of the experiment.

So, the final conclusion was that keystroke dynamics would greatly increase the security of a password authentication mechanism, but more work is required, which will lead to future studies, in order to make the mechanism more trustworthy, so it can be used commercially.

Glossary

In the table that follows, terms that were used in the thesis and the experiment are explained:

API	Application Programming Interface
BEM	Biometric Evaluation Methodology
CER	Crossover Error Rate
DNA	Deoxyribonucleic acid
EER	Equal Error Rate
EM	Expectation Maximization
EPIC	Electronic Privacy Information Center
EPSRC	Engineering and Physical Sciences Research Council
FAR	False Acceptance Rate
FER	Failure to Enroll Rate
FERET	Facial Recognition Technology
FMR	False Match Rate
FN	False Negative
FNMR	False Non-Match Rate
FRR	False Rejection Rate
FP	False Positive
GAR	Genuine Accept Rate

GMM	Gaussian Mixture Model
HMM	Hidden Markov Models
OAE	Otoacoustic Emission
ROC	Receiver Operating Characteristic
SVM	Support Vector Machines
TN	True Negative
TP	True Positive
URL	Unified Resource Locator

References

[1] A. K. Jain, A. Ross, and S. Prabhakar , "An introduction to biometric recognition", IEEE Transactions on Circuits and Systems for Video Technology, 2004.

[2] A. Bromme, A classification of biometric signatures, Multimedia and Expo, 2003, ICME '03, Proceedings. 2003 International Conference on, 2003.

[3] The Biometric Consortium, Introduction to Biometrics, accessed at 27Th September 2012. available at: http://www.biometrics.org/introduction.php

[4] EPSRC, Otoacoustic Emission Based Biometric Systems, accessed at 2009. available at: http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=EP/E015522/1

[5] M. Hildebrandt , Privacy and Identity, 2005, in: E. Claes and A. Duff (ed.), Privacy and the Criminal Law, proceedings of the Conference on Privacy and the Criminal Law, 2004.

[6] P. Ricoeur , Oneself as another, Chicago and London: University of Chicago Press 1992.

[7] J. Claessens, C. Díaz, S. Nikova, B. De Win, C. Goemans, M. Loncke, V. Naessens, S. Seys, B. De Decker, J. Dumortier, and B. Preneel, "Applications Requirements for Controlled Anonymity," APES Deliverable D7, 2003.

[8] A. Durand, Three Tiers of Identity, Digital Identity World, accessed at 2002.

available at : http://www.digitalidworld.com/print.php?sid=26

[9] M. Hildebrandt, and J. Backhouse eds., "Descriptive analysis and inventory of profiling practices", FIDIS deliverable 7.2, 2005

[10] J. Renaghan, "Etched in Stone," Zoogoer, August 1997, Smithsonian National Zoological Park, 2005.

[11] "Dermatoglyphics," Hand Analysis, International Institute of Hand Analysis, 2005.

[12] Z. McMahon, Biometrics: History, Indiana University, Indiana University Computer Science Department, accessed at 2005.

available at http://www.cs.indiana.edu/~zmcmahon/biometrics-history.htm

[13] A. Bertillon, accessed at 27th September 2012.

available at: http://en.wikipedia.org/wiki/Bertillon

[14] K. Bonsor, "How Facial Recognition Systems Work", accessed at 27th September 2012.

available at http://electronics.howstuffworks.com/facial-recognition.htm

[15] K. S. Killourhy, and R. A. Maxion, Comparing Anomaly-Detection Algorithms for Keystroke Dynamics, in IEEE/IFIP International Conference on Dependable Systems and Networks (DSN-2009), (June 29–July 2, 2009, Estoril, Lisbon, Portugal), IEEE Computer Society, Los Alamitos, CA, 2009, pp. 125–134.

[16] I. von Hardenberg, "Warum Neugeborene mehr wissen, als Große manchmal ahnen", GEO (7), Hamburg, 2001, pp.27-42.

[17] M. Amberger, S. Fischer, and J. Rößler, "Biometrische Verfahren – Studie zu State of the Art", Erlangen, accessed at 2003.

available at: http://www.wi3.unierlangen.de/forschung/Biometrie_StateOfTheArt/Biometrie.pdf

[18] M. Gasson, M. Meints, and K. Warwick, “D3.2: A study on PKI and biometrics”, accessed at 2005. available at: http://www.fidis.net/resources/deliverables/hightechid/#c1785

[19] J. G. Webster ed., Medical Instrumentation, 2nd ed. Boston: Houghton Mifflin Company, 1992, Chapter 5, pp. 249-250.

[20] W. Hirsch, and J.U. Schweichel, “Morphological Evidence Concerning the Problem of Skin Ridge Formation,” Journal of Mental Deficiency Research, 1973, Vol. 17, pp. 58-72.

[21] E. Locard, “Les Pores et L’Identification des Criminels,” Biologica, revue scientifique de medecin, Vol. 22, 1912, pp. 357-362.

[22] D. Hammelev, D. Madden, S. Nørby, and J. Turner, DNA profiling, European Initiative for Biotechnology Education (EIBE), 1998.

[23] M. A. Jobling, and P. Gill, “Encoded evidence: DNA in forensic analysis”, Nature (5), 2004, pp.739-52.

[24] M. Benecke, “Genetischer Fingerabdruck in Enzyklopädie Naturwissenschaft und Technik”, 2. Auflage, 6. Ergänzungslieferung, ecomed Verlagsgesellschaft AG & Co. KG, Landsberg, accessed at 2001. available at: http://www.benecke.com/dna.pdf

[25] R. Brunelli, and T. Poggio, “Face Recognition: Features versus Templates,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 10, 1993, pp.1042-1052.

[26] M.-H. Yang, D. J. Kriegman and N. Ahuja, "Detecting Faces in Images: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, 2002, pp. 34-58,.

[27] K. M. Lam, and H. Yan, “An Analytic-to-Holistic Approach for Face Recognition based on a Single Frontal View,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 20, No. 7, 1998, pp.673-686.

[28] A. S. Tolba, A. H. El-Baz, and A. A. El-Harby, “Face Recognition – A Literature Review”, International Journal of Signal Processing, Vol.2, No.2, 2005, pp.88-103.

[29] K. S. Killourhy, “A Scientific Understanding of Keystroke Dynamics”, School of Computer Science, Computer Science Department, Carnegie Mellon University Pittsburgh, PA 15213, 2012.

[30] G. Forsen, M. Nelson, and J. Raymond Staron, Personal attributes authentication techniques. Technical Report RADC-TR-77-333, Rome Air Development Center, October 1977.

[31] R. S. Gaines,W. Lisowski, S. J. Press, and N. Shapiro, Authentication by keystroke timing: Some preliminary results. Technical Report R-2526-NSF, RAND Corporation, 1980.

[32] R. Joyce and G. Gupta, Identity authentication based on keystroke latencies. *Communications of the ACM*, Vol. 33, No. 2, 1990, pp. 168–176.

[33] L. C. F. Araújo, L. H. R. Sucupira Jr., M. G. Lizárraga, L. L. Ling, and J. B. T. Yabuuti, User authentication through typing biometrics features. *IEEE Transactions on Signal Processing*, Vol. 53, No. 2, 2005, pp. 851–855.

[34] M. Brown, and S. J. Rogers, User identification via keystroke characteristics of typed names using neural networks, *International Journal of Man-Machine Studies*, Vol. 39, No. 6, 1993, pp. 999–1014.

[35] M. S. Obaidat, and B. Sadoun, Verification of computer users using keystroke dynamics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 27, No. 2, 1997, pp. 261–269.

[36] S. Cho, C. Han, D. H. Han, and H.-I. Kim, Web-based keystroke dynamics identity verification using neural network, *Journal of Organizational Computing and Electronic Commerce*, Vol. 10, No. 4, 2000, pp. 295–307.

[37] B. Hussien, R. McLaren, and S. Bleha, An application of fuzzy algorithms in a computer access security system, *Pattern Recognition Letters*, Vol. 9, No. 1, 1989, pp. 39–43.

[38] B. Hussien, R. McLaren, and S. Bleha, An application of fuzzy algorithms in a computer access security system, *Pattern Recognition Letters*, Vol. 9, No. 1, 1989, pp. 39–43.

[39] W. G. de Ru, and J. H. P. Eloff, Enhanced password authentication through fuzzy logic, *IEEE Expert*, Vol. 12, No. 6, 1997, pp. 38–45.

[40] S. Haider, A. Abbas, and A. K. Zaidi, A multi-technique approach for user identification through keystroke dynamics (SMC 2000), In *IEEE International Conference on Systems, Man and Cybernetics*, IEEE, Piscataway, NJ, Nashville, Vol. 2, 2000, pp. 1336–1341.

[41] S. Mandujano, and R. Soto, Deterring password sharing: User authentication via fuzzy c-means clustering applied to keystroke biometric data, In *5th Mexican International Conference on Computer Science*, Colima, Mexico, IEEE Computer Society, Los Alamitos, CA, 2004, pp. 181–187

[42] D. Tran, W. Ma, G. Chetty, and D. Sharma, Fuzzy and markov models for keystroke biometrics authentication, WSEAS Press, In *Proceedings of the 7th WSEAS International Conference on Simulation, Modelling and Optimization*, Beijing, China, 2007, pp. 89–94.

[43] E. Yu, and S. Cho, GA-SVM wrapper approach for feature subset selection in keystroke dynamics identity verification, In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, IEEE, Piscataway, NJ, Portland, OR, Vol. 3, pp. 2253–2257, 2003.

[44] K.-S. Sung, and S. Cho, GA SVM wrapper ensemble for keystroke dynamics authentication, In *Advances in Biometrics (ICB 2006)*, Vol. 3832 of *Lecture Notes in Computer Science (LNCS)*, pp. 654–660, January 5–7, 2006, Hong Kong, China, 2005. Springer-Verlag, Berlin.

[45] C. C. Loy, W. K. Lai, and C. P. Lim, Keystroke patterns classification using the ARTMAPFD neural network, In *3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, IEEE, Piscataway, NJ, Kaohsiung, Taiwan, Vol. 1, 2007, pp. 61-64.

[46] R. Giot, M. El-Abed, and C. Rosenberger, Keystroke dynamics with low constraints SVM based passphrase enrollment, In *IEEE Third International Conference on Biometrics: Theory, Applications and Systems* , Washington, DC, 2009, pp. 425–430.

[47] N. Clarke, S. Furnell, and P. Reynolds, "Biometric Authentication for Mobile Devices," in Proceedings of the 3rd Australian Information Warfare and Security Conference 2002, 2002, pp. 61–69.

[48] J. Ilonen, "Keystroke dynamics" Lappeenranta University of Technology, Skinnarilankatu 34, 53850 Lappeenranta, Finland.

[49] S. Memon, N. Manivannan, and W. Balachandran, "Active pore detection for liveness in fingerprint identification system", Telecommunications Forum (TELFOR), 2011.

[50] K. A. Nixon, V. Aimale, R. K. Rowe, Spoof Detection Schemes, Handbook of Biometrics, Springer, 2007.

Ακολουθεί μια εκτεταμένη περίληψη του παραπάνω κειμένου στα ελληνικά.

Χαροκόπειο Πανεπιστήμιο
Τμήμα Πληροφορικής και Τηλεματικής

Μελέτη και ανάπτυξη μηχανισμών για την επιβεβαίωση ταυτότητας μέσω των δυναμικών πληκτρολόγησης

Πτυχιακή εργασία του

Αβραμίδη Λουκά

Τριμελής Επιτροπή:
Δημήτριος Μιχαήλ
Βασσιλική Ανδρόνικου
Ηρακλής Βαρλάμης

Αθήνα, Σεπτέμβριος 2012

Πίνακας Περιεχομένων

Πρόλογος v

1 Εισαγωγή 1

1.1 Ένοια Βιομετρικών Συστημάτων 1

1.2 Όραμα και Ανοιχτά Τεχνολογικά Θέματα 2

1.3 Συνεισφορά – Καινοτομία Πτυχιακής Εργασίας 3

1.4 Οργάνωση Διατριβής 4

2 Βιομετρικά Συστήματα 6

2.1 Η Έννοια της Διαχείρισης Ταυτότητας 6

2.2 Χαρακτηριστικά προς Ταυτοποίηση 7

2.3 Ιστορική Ανασκόπηση Βιομετρικής Ταυτοποίησης και Επιβεβαίωσης Ταυτότητας 8

2.4 Ποιοτικά Χαρακτηριστικά των Βιομετρικών Γνωρισμάτων 10

2.5 Κατηγοριοποιήσεις Βιομετρικών Χαρακτηριστικών και Συστημάτων 11

2.6 Παραδείγματα Βιομετρικών Χαρακτηριστικών 14

2.6.1 Πρόσωπο 14

2.6.2 Δαχτυλικά Αποτυπώματα 17

2.6.3 Δυναμικές της Πληκτρολόγισης 19

2.7 Βιομετρικές εφαρμογές 20

2.8 Οι δύο φάσεις λειτουργίας των Βιομετρικών Συστημάτων 21

2.9 Είδη Σφαλμάτων Βιομετρικών Συστημάτω 22

Έννοιες 25

Βιβλιογραφία 27

Πίνακας Εικόνων

Εικόνα 1 Χαρακτηριστικά προς ταυτοποίηση 8
Εικόνα 2 Κατηγορίες Βιομετρικών Χαρακτηριστικών 13
Εικόνα 3 Επτά κοινά πρότυπα που χρησιμοποιούνται από το FBI [12] 18
Εικόνα 4 Χρόνοι αναμονής μεταξύ πληκτρολογήσεις όταν γράφετε τη λέξη "password" από τρεις διαφορετικούς ανθρώπους. 20
Εικόνα 5 Γενικά Στάδια Βιομετρικού Συστήματος 22
Εικόνα 6 Σχέσεις μεταξύ FAR FRR EER 24

Πρόλογος

Οι δυναμικές της πληκτρολόγησης αποτελούν ένα βιομετρικό χαρακτηριστικό που ερευνάτε από τότε που ξεκίνησαν να χρησιμοποιούνται τα πληκτρολόγια για την γραφή στους υπολογιστές. Είναι μια πολύ χρήσιμος, φτηνός στην εφαρμογή και μεταφέρσιμος τρόπος για την πιστοποίηση χρήστη μέσω της εισαγωγής συνθηματικού. Ένα συνθηματικό είναι εύκολο να μαθευτεί από λάθος άτομα, αλλά με έναν τέτοιον μηχανισμό πιστοποίησης, η ασφάλεια του συστήματος μπορεί να αυξηθεί κατά πολύ.

Το παρακάτω κείμενο αποτελεί μια εκτεταμένη περίληψη της έρευνας που έγινα αρχικά στα Αγγλικά, και περιέχει τα βασικά στοιχεία για την κατανόηση των βιομετρικών συστημάτων, και των δυναμικών πληκτρολόγησης.

1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται η έννοια των βιομετρικών συστημάτων και στη συνέχεια αναλύεται η διάρθρωση της παρούσης Διατριβής και περιγράφεται συνοπτικά το περιεχόμενο των κεφαλαίων.

1.1 Ένοια Βιομετρικών Συστημάτων

Οι βιομετρικές τεχνολογίες έχουν προκαλέσει πλήθος συζητήσεων και διαφωνιών τα τελευταία χρόνια καθώς τόσο οι κυβερνήσεις όσο και ο επιχειρηματικός κόσμος σταδιακά δείχνουν αυξανόμενο ενδιαφέρον στην υιοθέτηση τεχνολογιών βασισμένων σε βιομετρικά χαρακτηριστικά σε μεγάλης κλίμακας εφαρμογές σχετικές με την ταυτοποίηση και πιστοποίηση της ταυτότητας των πολιτών, όπως τα βιομετρικά διαβατήρια. Το κυβερνητικό και επιχειρηματικό αυτό ενδιαφέρον έχει ενδυναμωθεί από το αυξανόμενο εμπορικό ενδιαφέρον για τις βιομετρικές εφαρμογές και τις δυνατότητες αυτών καθώς και από την πυρετώδη ερευνητική δραστηριότητα στον εν λόγω χώρο η οποία τα τελευταία χρόνια έχει προσφέρει σημαντική βελτίωση στις επιδόσεις των συστημάτων Η χρήση των βιομετρικών χαρακτηριστικών για ταυτοποίηση και πιστοποίηση της ταυτότητας ενός ατόμου δεν αποτελεί όμως κάποια

επαναστατική μέθοδο των τελευταίων ετών, αλλά βασίζεται στην ίδια τη λειτουργία του ανθρώπου να αναγνωρίζει άτομα από το πρόσωπο, τη φωνή, το περπάτημα.

Ο όρος *βιομετρικός* (biometric : bio + metric) προέρχεται από τον συνδυασμό των ελληνικών λέξεων «*βίος*» που σημαίνει ζωή και «*μετρικός*» που σημαίνει αυτός που μετράει. Έτσι, οι βιομετρικές τεχνολογίες βασίζονται σε ανθρώπινα χαρακτηριστικά με συγκεκριμένες ιδιότητες όπως θα δούμε και στο Κεφάλαιο 3 της παρούσης Διατριβής τα οποία επιτρέπουν με κατάλληλη λήψη, αποτύπωση, ανάλυση και επεξεργασία την ταυτοποίηση του ατόμου στο οποίο ανήκουν.

Σύμφωνα με τους Jain et al [1]ένα βιομετρικό σύστημα αποτελεί επί της ουσίας ένα σύστημα ταυτοποίηση προτύπων το οποίο λειτουργεί με την απόκτηση βιομετρικών δεδομένων από ένα άτομο, την εξαγωγή ενός συνόλου χαρακτηριστικών γνωρισμάτων από τα δεδομένα αυτά, και τη σύγκριση αυτού του συνόλου με ένα προαποθηκευμένο σύνολο πρότυπων χαρακτηριστικών γνωρισμάτων.

Ο Bromme [2] θέλοντας να αποτυπώσει την έννοια των *βιομετρικών συστημάτων στον χώρο της ασφάλειας των πληροφοριακών συστημάτων* αναφέρει ότι αποτελούν μια μέθοδο ταυτοποίηση του ατόμου βασισμένη στην απόκτηση ανθρώπινων βιολογικών χαρακτηριστικών μέσω αισθητήρων, στη μέτρηση των βιομετρικών χαρακτηριστικών (ανεπεξέργαστα δεδομένα και δεδομένα βαθμονόμησης αισθητήρα), στον υπολογισμό των βιομετρικών δεδομένων και των βιομετρικών προτύπων, και στην πιστοποίηση και ταυτοποίηση με βάση βιομετρικά πρότυπα σε σχέση με μαθηματικούς ορισμούς των μετρικών και των μετρικών χώρων.

Το Biometric Consortium [3] τα ορίζει ως αυτοματοποιημένες μεθόδους ταυτοποίησης ενός ατόμου με βάση ένα φυσιολογικό και συμπεριφορικό χαρακτηριστικό. Μεταξύ αυτών των μετρηθέντων χαρακτηριστικών είναι το πρόσωπο, τα δακτυλικά αποτυπώματα, η γεωμετρία του χεριού, η γραφή, η ίρις και η φωνή.

1.2 Όραμα και Ανοιχτά Τεχνολογικά Θέματα

In the past few years, research in biometric systems has advanced rapidly all over the world, with a vast number of researchers heavily working on advancements in different fields such as

technology, law, social sciences and economics. Especially in the technological area, specialists work in a variety of domains in order to advance or build new algorithms for increasing the speed, the accuracy and the reliability of those systems. At the same time, by taking into consideration the great variety of human characteristics, researchers look for new biometric features that can be used to increase the accuracy of identification/verification systems and to reduce the cost of their application and their intrusiveness in human life. For example, a rather new biometric feature has been researched upon the past years, the otoacoustic emissions(OAEs), which is based on sounds that the ear produces when it is stimulated. According to the Engineering and Physical Sciences Research Council (EPSRC) [4] those emissions are created by the ability of the human body to boost very small sounds and according to their researches, those emissions are unique for every person. They are produced by the spiral screw, an organ in the human ear.

At the same time, great efforts are being made to apply biometrics for the advancement of security systems. In consideration of the required processing power and the privacy and information security issues, which are bound with the use of biometric systems, the scientific interest has befallen on the use of the already existing tools and their advancement, so they can support the computing requirements of any biometric applications.

1.3 Συνεισφορά – Καινοτομία Πτυχιακής Εργασίας

Τα τελευταία χρόνια η έρευνα στο πεδίο των βιομετρικών συστημάτων προχωρά με γοργούς ρυθμούς τόσο σε Ευρωπαϊκό όσο και σε παγκόσμιο επίπεδο με πλήθος ερευνητών από διαφορετικούς χώρους, όπως τεχνολογικό, νομικό, κοινωνικό, οικονομικό και κυβερνητικό. Στα πλαίσια αυτής της πυρετώδους ερευνητικής δραστηριότητας και του διαρκώς διευρυνόμενου πεδίου εφαρμογής των βιομετρικών συστημάτων, ομάδες έμπειρων επιστημόνων και ερευνητών έχουν διατυπώσει και εργάζονται αναφορικά με το όραμα και τα ανοιχτά τεχνολογικά θέματα των βιομετρικών τεχνολογιών.

Δεδομένων των διαφορετικών τεχνολογιών που βρίσκονται υπό την ομπρέλα των βιομετρικών συστημάτων, όπως είναι η ταυτοποίηση προσώπου, φωνής, περπατήματος, τρόπου υπογραφής, οι ειδικοί σε κάθε χώρο εργάζονται εντατικά προκειμένου να αυξήσουν

την αποτελεσματικότητα των αλγορίθμων ή να αναπτύξουν νέους πιο αποδοτικούς τόσο όσον αφορά στις επεξεργαστικές απαιτήσεις όσο και στην ακρίβεια των αποτελεσμάτων αυτών. Παράλληλα, δεδομένου του πλούτου των βιομετρικών χαρακτηριστικών του ανθρώπου έρευνα πραγματοποιείται προκειμένου να προσδιοριστούν νέα βιομετρικά χαρακτηριστικά επιζητώντας τόσο την ενίσχυση της ακρίβειας των βιομετρικών συστημάτων όσο και τη μείωση του κόστους αυτών καθώς και της επεμβατικότητάς τους στην ανθρώπινη ζωή. Έτσι, για παράδειγμα, ένα νέο βιομετρικό χαρακτηριστικό το οποίο είναι υπό έρευνα αυτήν την περίοδο είναι οι ωτοακουστικές εκπομπές (Otoacoustic Emissions - OAEs), οι οποίες βασίζονται στους ήχους που παράγει το αυτί ανταποκρινόμενο σε ακουστικό ερέθισμα. Σύμφωνα με το Βρετανικό Ερευνητικό Συμβούλιο Μηχανικής και Φυσικών Επιστημών (Engineering and Physical Sciences Research Council (EPSRC)) [4] η πηγή αυτών των εκπομπών έγκειται στη λειτουργία του ανθρώπινου σώματος να ενισχύει ήχους χαμηλού επιπέδου και σύμφωνα με τις μελέτες τους οι εκπομπές αυτές είναι μοναδικές ανάμεσα στους ανθρώπους. Οι εκπομπές αυτές είναι ήχοι παραγόμενοι στο ανθρώπινο αυτί οι οποίοι πηγάζουν από τον σπειροειδή κοχλία στο εσωτερικό του αυτιού.

Παράλληλα και στα πλαίσια της υιοθέτησης των βιομετρικών συστημάτων για τη βελτίωση συστημάτων ασφαλείας, σημαντικές προσπάθειες λαμβάνουν χώρα για την ανάπτυξη τέτοιων συστημάτων πραγματικού χρόνου. Δεδομένης της απαιτούμενης υπολογιστικής ισχύος αλλά και των θεμάτων ιδιωτικότητας και ασφάλειας των δεδομένων που είναι άρρηκτα συνδεδεμένα με τη χρήση των βιομετρικών συστημάτων σε εφαρμογές μεγάλης κλίμακας, το ερευνητικό ενδιαφέρον έχει στραφεί προς την εκμετάλλευση υπαρχουσών υποδομών καθώς και την περαιτέρω ανάπτυξη αυτών προκειμένου να είναι σε θέση να υποστηρίξουν τις αυξημένες λειτουργικές και μη απαιτήσεις των βιομετρικών εφαρμογών.

1.4 Οργάνωση Διατριβής

Chapter 1 Εισαγωγή

Εισαγωγή του όρου "βιομετρικού συστήματος". Μερικά πράγματα που είπε σχετικά με πρόσφατα ευρήματα σε αυτόν τον τομέα. Ο τρόπος με βιομετρικά χαρακτηριστικά μπορεί να εφαρμοστεί σε συστήματα ασφαλείας εξηγείται, και είναι δυναμική

πληκτρολόγησης παρουσιάζεται ως μια μέθοδο ελέγχου ταυτότητας παρόμοια με την υπογραφή χειρογράφου.

Chapter 2 Βιομετρικά Συστήματα

Μια λίγο περισσότερο σχετικά με τα βιομετρικά συστήματα. Μιλάμε για τη σημασία της διαχείρισης ταυτότητας και παρουσιάζουμε διάφορα χαρακτηριστικά που μπορούν να χρησιμοποιηθούν για τη βιομετρική ταυτοποίηση και επαλήθευση της ταυτότητας. Περνάμε μέσα από μια σύντομη ιστορική αναδρομή για το πώς χρησιμοποιούνται τα βιομετρικά χαρακτηριστικά 30000BC, μέχρι σήμερα. Επιπλέον, μιλάμε για τις ιδιότητες των βιομετρικών στοιχείων, καθώς και σχετικά με τα διάφορα είδη των βιομετρικών στοιχείων και συστημάτων. Λέμε επίσης λίγα λόγια για τα λάθη που μπορεί να συναντήσετε κατά τη χρήση συστημάτων όπως αυτό.

2 Βιομετρικά Συστήματα

Στο κεφάλαιο αυτό, ο όρος και η σημασία της διαχείρισης της ταυτότητας εισάγεται, τα είδη των βιομετρικών συστημάτων που υπάρχουν, οι φάσεις στις οποίες εργάζονται, τα διάφορα είδη των χαρακτηριστικών που μπορεί να εξαχθεί και να χρησιμοποιηθεί για την επικύρωση ή την ταυτοποίηση ορισμένων μεμονωμένων, και οι τα είδη των λαθών που θα μπορούσαν να προκύψουν κατά τη χρήση ενός βιομετρικού συστήματος.

2.1 Η Έννοια της Διαχείρισης Ταυτότητας

Η *διαχείριση ταυτότητας* αφορά στη διαχείριση του κύκλου ζωής της ταυτότητας ενός υποκειμένου ή ενός αντικειμένου. Ο όρος αυτός εμπεριέχει πολλά διαφορετικά ζητήματα σχετικά με την ταυτότητα, όπως είναι η εξακρίβωσή της (διαδικασία ταυτοποίησης χρήστη) αλλά και θέματα που έχουν να κάνουν με τη διαφύλαξη, την εμπιστευτικότητα, την ασφάλεια, την πρόσβαση, τη διαλειτουργικότητα. Ωστόσο, η διαχείριση της ταυτότητας είναι ένα ζήτημα πιο ευρύ και οριζόντιο που ανακύπτει συνολικότερα και έχει να κάνει με το σύνολο των πληροφοριών που καθορίζουν μοναδικά μια οντότητα σε ένα πολλαπλό και ευρύ φάσμα δραστηριοτήτων με τρόπο που να είναι άμεσα αναγνωρίσιμη.

Στον επιχειρηματικό και τεχνολογικό κόσμο ο όρος αυτός αφορά σε ένα σύνολο διαδικασιών, μεθόδων και τεχνολογιών που διευκολύνουν και ελέγχουν την πρόσβαση χρηστών σε εφαρμογές και πόρους με παράλληλη προστασία εμπιστευτικής προσωπικής και επιχειρησιακής πληροφορίας από μη εξουσιοδοτημένους χρήστες.

Στον επιχειρηματικό κόσμο συγκεκριμένα η διαχείριση ταυτότητας συνιστά σύγκλιση τεχνολογικών και επιχειρησιακών διαδικασιών, η οποία στοχεύει στο να:

1 Καταστήσει δυνατό ένα υψηλότερο επίπεδο ηλεκτρονικής επιχειρηματικότητας με την επιτάχυνση της μετάβασης σε ένα συνεπές σύνολο από πρότυπα διαχείρισης ταυτότητας

2 Μειώσει την πολυπλοκότητα της ενοποίησης των επιχειρηματικών εφαρμογών

3 Διαχειριστεί τη ροή χρηστών που εισέρχονται, χρησιμοποιούν και εγκαταλείπουν τον οργανισμό

4 Υποστηρίξει σφαρικές προσεγγίσεις και σχήματα για συγκεκριμένες κατηγορίες λειτουργικών έργων

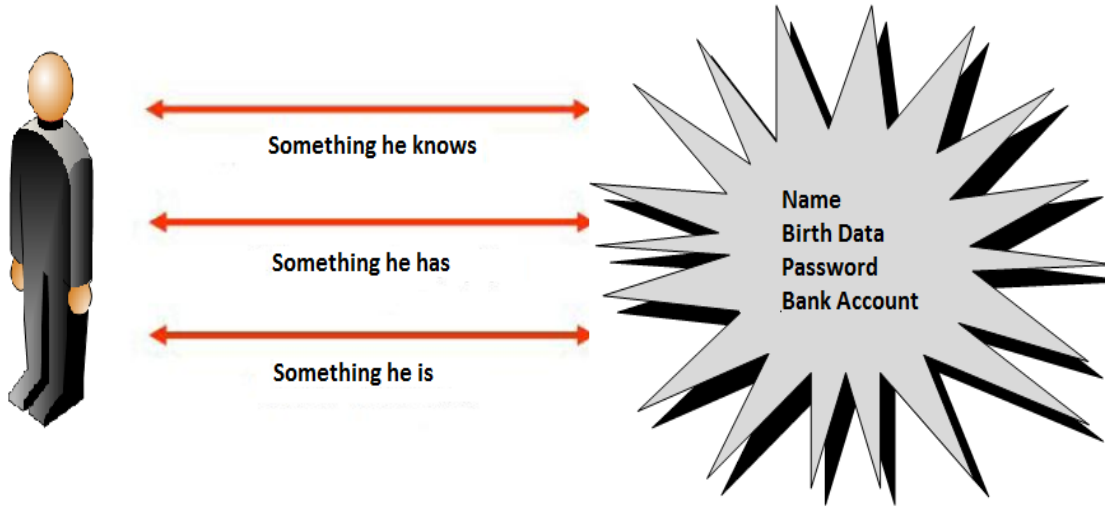
Ανταποκρίνεται στην πίεση από τον διαρκώς αυξανόμενο αριθμό διαδικτυακών επιχειρηματικών εφαρμογών που χρειάζονται μεγαλύτερη ενοποίηση για δραστηριότητες όπως το single sign-on.

2.2 Χαρακτηριστικά προς Ταυτοποίηση

Η διαδικασία ταυτοποίησης ενός ατόμου μπορεί να βασίζεται σε πληθώρα χαρακτηριστικών του τα οποία γενικώς διακρίνονται σε :

- Κάτι που το άτομο γνωρίζει: αφορά σε πληροφορία την οποία το άτομο γνωρίζει, όπως είναι ο προσωπικός του κωδικός για την πρόσβαση σε ένα σύστημα,
- Κατι που το άτομο έχει: περιλαμβάνει περιπτώσεις κατά τις οποίες το άτομο είναι κάτοχος ενός αντικειμένου αναγνωριστικού, όπως είναι μια κάρτα, ψηφιακά πιστοποιητικά, κλπ,

- Κάτι που το άτομο είναι ή κάνει: σχετίζεται με γνωρίσματα με τα οποία το άτομο είναι άμεσα συνδεδεμένα, με χαρακτηριστικό παράδειγμα τα βιομετρικά του χαρακτηριστικά.



Εικόνα 1 Χαρακτηριστικά προς ταυτοποίηση

Όλα αυτά τα χαρακτηριστικά διαφέρουν ως προς την πρακτικότητα τους, το κόστος εφαρμογής τους, το πεδίο εφαρμογής τους, την αξιοπιστία τους, την αποτελεσματικότητα τους, την ασφάλειά τους αλλά και τον σεβασμό τους προς την ιδιωτικότητα του ατόμου.

2.3 Ιστορική Ανασκόπηση Βιομετρικής Ταυτοποίησης και Επιβεβαίωσης Ταυτότητας

Η χρήση των βιομετρικών συστημάτων δεν αποτελεί κάποια νέα μέθοδο ταυτοποίησης και επιβεβαίωσης της ταυτότητας ατόμων. Έτσι, από τις απαρχές του πολιτισμού οι άνθρωποι βασίζονταν στο ανθρώπινο πρόσωπο προκειμένου να διακρίνουν γνωστά και μη άτομα. Με τα χρόνια η διαδικασία αυτή έγινε πιο πολύπλοκη καθώς ο πληθυσμός αυξανόταν και καθώς η εφέυρεση νέων μεθόδων μεταφοράς έφερνε νέα άτομα στις κάποτε μικρές κοινωνίες.

Στο πέρασμα των αιώνων πληθώρα ανθρώπινων χαρακτηριστικών χρησιμοποιήθηκαν ως μια

πιο επίσημη μορφή ταυτοποίησης. Συγκεκριμένα, σε μια σπηλιά η οποία χρονολογείται ότι είναι παραπάνω από 31000 ετών, βρέθηκαν τοιχογραφίες οι οποίες εκτιμάται ότι δημιουργήθηκαν από προϊστορικούς άνθρωπους οι οποίοι έμεναν εκεί. Γύρω από τις τοιχογραφίες αυτές βρέθηκαν διάφορα αποτυπώματα ανθρώπινης παλάμης τα οποία θεωρείται ότι αποτελούσαν «υπογραφή» του δημιουργού τους [10]. Στοιχεία επίσης συνηγορούν ότι η χρήση των αποτυπωμάτων ως μέσο ταυτοποίησης χρονολογείται στα 500 π.Χ. όταν οι Βαβυλώνιοι αποτύπωναν τις συναλλαγές τους πάνω σε πλάκες από πηλό οι οποίες συμπεριελάμβαναν δακτυλικά αποτυπώματα [11]. Ο Ισπανός εξερευνητής και συγγραφέας Joao de Barros κατέγραψε μεταξύ των άλλων ότι οι παρόμοια χρήση των δαχτυλικών αποτυπωμάτων απαντάται και στην Κίνα του 14^{ου} αιώνα οπότε και οι έμποροι τα χρησιμοποιούσαν κατά τις εμπορικές τους συναλλαγές. Μεταξύ των άλλων βασίζονταν σε αυτό το βιομετρικό χαρακτηριστικό καθώς και το αποτύπωμα του ποδιού τα οποία ελάμβαναν μέσω μελανιού για να ξεχωρίζουν τα παιδιά [12]. Σύμφωνα με τον ίδιο συγγραφέα, οι αρχαίοι Αιγύπτιοι έμποροι βασίζονταν σε φυσικούς περιγραφείς για τον διαχωρισμό των έμπιστων εμπόρων γνωστής φήμης και με προγενέστερες επιτυχείς συναλλαγές από νέους εμπόρους.

Στα μεταγενέστερα χρόνια η χρήση βιομετρικών χαρακτηριστικών παρατηρήθηκε πιο έντονα. Το 1890 ο Alphonse Bertillon, ένας Γάλλος αστυνόμος μελέτησε τη μηχανική του σώματος καθώς και μετρήσεις προκειμένου να καταστεί δυνατή η σύλληψη εγκληματιών [13]. Η αστυνομία βασίστηκε στη μέθοδο αυτή, την επονομαζόμενη Bertillonage μέθοδο, μέχρι τη στιγμή που αναγνώρισε λανθασμένα κάποιους ανθρώπους. Η μέθοδος αυτή γρήγορα εγκαταλείφθηκε, ενώ ξαναεισήχθει από τον Richard Edward Henry της Scotland Yard.

Στις αρχές του 20^{ου} αιώνα ο Karl Pearson, ένας ερευνητής εφαρμοσμένων μαθηματικών, μελέτησε τη βιομετρία στο University College of London. Κατά τη διάρκεια της έρευνάς του οδηγήθηκε σε σημαντικές ανακαλύψεις στον χώρο αυτό μέσω της μελέτης στατιστικής ιστορίας και συσχέτισης, την οποία και εφάρμοσε στην εξέλιξη των ζώων. Το έργο του αυτό περιελάμβανε τη μέθοδο των στιγμών, το Pearson σύστημα των καμπύων, συσχέτιση και την chi-squared δοκιμή.

Αργότερα, κατά τις δεκαετίες 1960 και 1970, διαδικασίες βιομετρικής πιστοποίησης μέσω της υπογραφής αναπτύχθηκαν, αλλά ο βιομετρικός χώρος έμεινε σταθερός μέχρι τη στιγμή που

τα στρατιωτικά γραφεία και τα γραφεία ασφαλείας προχώρησαν στην έρευνα και στην ανάπτυξη της βιομετρικής τεχνολογίας πέραν των δαχτυλικών αποτυπωμάτων.

Τα τελευταία χρόνια και με διαρκώς αυξανόμενη την προσπάθεια για βελτίωση των συστημάτων ταυτοποίησης και επιβεβαιώσης ταυτότητας τα οποία κυρίως βασίζονταν σε κωδικούς και κάρτες, συνοδευόμενα από σειρά προβλημάτων τα οποία αναλύθηκαν στο προηγούμενο κεφάλαιο, παρατηρείται μια έντονη δραστηριότητα στον χώρο των βιομετρικών συστημάτων. Μια σημαντική εφαρμογή αυτών αποτέλεσε το Super Bowl στην Tampa, Florida, το 2001 [14] κατά το οποίο έγινε χρήση συστήματος αναγνώρισης προσώπου με κάθε εικόνα προσώπου από τους 100000 θεατές να έχει καταγραφεί από κάμερες ασφαλείας και να έχει ελεγχθεί ηλεκτρονικά σε σχέση με εικόνες κακοποιών από την αστυνομία της Tampa. Αυτή η κίνηση συνδυαζόμενη από τη μη επιτυχή λειτουργία του συστήματος προκάλεσε πληθώρα αντιδράσεων από υποστηρικτές των κοινωνικών ελευθεριών οι οποίοι κατήγγειλαν τότε τις βιομετρικές τεχνολογίες ως εχθρούς της ελευθερίας του ατόμου.

Παρ’όλα αυτά και ειδικά μετά το τρομοκρατικό κτύπημα της 11^{ης} Σεπτεμβρίου 2001 στις ΗΠΑ, πολλές αρχές παγκοσμίως προχώρησαν στην εγκατάσταση βιομετρικών τεχνολογιών στα αεροδρόμια προκειμένου να εντοπίσουν υπόπτους για τρομοκρατία. Λίγα χρόνια αργότερα και μετά από πληθώρα βομβιστικών επιθέσεων στον υπόγειο σιδηρόδρομο, η Μεγάλη Βρετανία υιοθέτησε τεχνολογίες αναγνώρισης προσώπου καταλήγοντας σήμερα το Λονδίνο να έχει περισσότερες από 200000 κάμερες σε δημόσιους χώρους για λόγους ασφαλείας.

2.4 Ποιοτικά Χαρακτηριστικά των Βιομετρικών Γνωρισμάτων

Όπως έχει ήδη αναφερθεί, οι βιομετρικές τεχνολογίες για ταυτοποίηση, επιβεβαίωση ταυτότητας αλλά και δημιουργία προφίλ ενός ατόμου βασίζονται σε συγκεκριμένα ανθρώπινα χαρακτηριστικά. Σε αυτό το σημείο τίθεται το ερώτημα, οι μετρήσεις ποιων ανθρώπινων χαρακτηριστικών μπορούν να χρησιμοποιηθούν για τον σκοπό αυτόν. Οποιοδήποτε ανθρώπινο φυσιολογικό ή/και συμπεριφορικό χαρακτηριστικό μπορεί να χρησιμοποιηθεί ως βιομετρικό χαρακτηριστικό αρκεί να ικανοποιεί τις ακόλουθες απαιτήσεις [1]:

- *Καθολικότητα (Universality)*: κάθε άτομο πρέπει να έχει αυτό το χαρακτηριστικό

- *Δυνατότητα Διαχωρισμού (Distinctiveness)*: κάθε δύο άτομα θα πρέπει να είναι επαρκώς διαφορετικά μεταξύ τους όσον αφορά σε αυτό το χαρακτηριστικό
- *Μονιμότητα (Permanence)*: το χαρακτηριστικό αυτό θα πρέπει να είναι επαρκώς αμετάβλητο (σε σχέση με το κριτήριο σύγκρισης) κατά τη διάρκεια μιας χρονικής περιόδου
- *Δυνατότητα συλλογής (Collectability)*: το χαρακτηριστικό αυτό θα πρέπει να μπορεί να μετρηθεί ποσοτικά

Εκτός όμως από τα παραπάνω, και λαμβάνοντας υπ’όψιν την πρακτική όψη του εγχειρήματος ανάπτυξης ενός βιομετρικού συστήματος, υπάρχουν κι άλλα θέματα που θα πρέπει να ληφθούν υπ’όψιν όπως:

- *Απόδοση (Performance)*: η οποία αφορά στη δυνατή προς επίτευξη ακρίβεια και ταχύτητα αναγνώρισης, αλλά και στους λειτουργικούς και περιβαλλοντολογικούς παράγοντες οι οποίοι επηρεάζουν την ακρίβεια και την ταχύτητα.
- *Δυνατότητα Αποδοχής (acceptability)*: η οποία δείχνει το βαθμό στον οποίο ποιοι άνθρωποι είναι πρόθυμοι να αποδεχτούν τη χρήση ενός συγκεκριμένου βιομετρικού αναγνωριστικού (identifier) στην καθημερινή τους ζωή,
- *Καταστρατήγηση (circumvention)*, η οποία αντανακλά το πόσο έυκολα μπορεί το σύστημα να ξεγελαστεί με τη χρήση δόλιων μεθόδων.

2.5 Κατηγοριοποιήσεις Βιομετρικών Χαρακτηριστικών και Συστημάτων

Ο πλούτος των βιομετρικών χαρακτηριστικών έχει οδηγήσει σε διάφορες κατηγοριοποιήσεις αυτών. Οι δύο κύριες κατηγορίες στις οποίες μπορούν να διακριθούν τα βιομετρικά χαρακτηριστικά είναι τα *φυσιολογικά* (ή *παθητικά*) και τα *συμπεριφορικά* (ή *ενεργητικά*). Τα φυσιολογικά βιομετρικά χαρακτηριστικά αφορούν σε ανθρώπινα χαρακτηριστικά τα οποία είναι εν γένει σταθερά ή αμετάβλητα όπως τα δαχτυλικά αποτυπώματα, η γεωμετρία του χεριού, η ίρις, η φωνή και το πρόσωπο. Τα συμπεριφορικά βιομετρικά χαρακτηριστικά

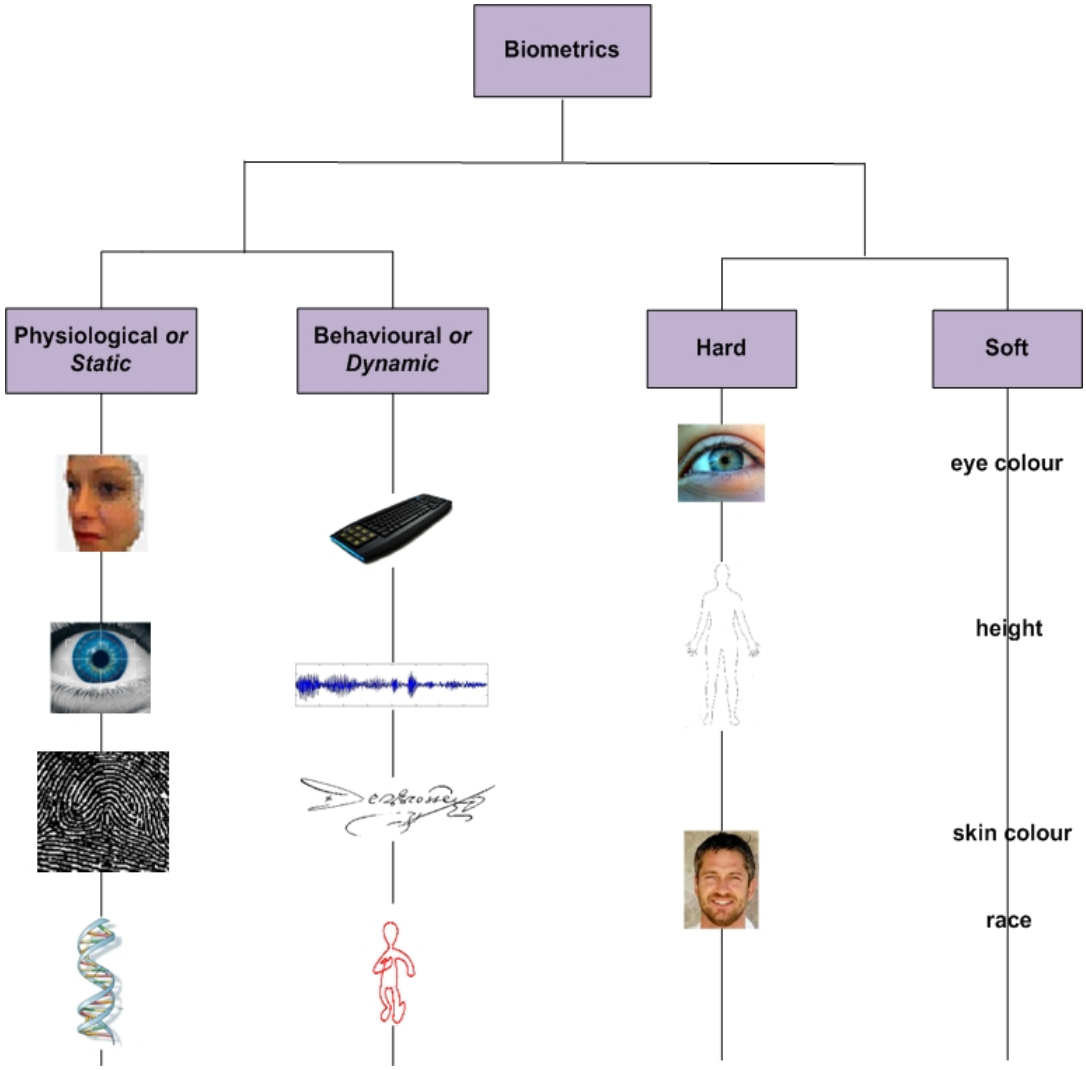
σχετίζονται με δεξιότητες ή λειτουργίες που πραγματοποιούνται από ένα άτομο μια δεδομένη στιγμή και για έναν συγκεκριμένο λόγο, όπως είναι για παράδειγμα η υπογραφή ή η πληκτρολόγηση κειμένου στον υπολογιστή.

Μια ακόμη κατηγοριοποίηση των βιομετρικών χαρακτηριστικών η οποία απαντάται είναι ο διαχωρισμός τους σε *στατικά* και *δυναμικά*. Ο όρος «στατικός» αναφέρεται στη μέτρηση ενός γνωρίσματος το οποίο δεν απαιτεί κάποια δράση κατά τη στιγμή της ταυτοποίησης ή της επιβεβαίωσης ταυτότητας. Με την έννοια του «δυναμικού» προσδιορίζονται μετρήσεις ενός γνωρίσματος κατά τη διάρκεια κάποιας δραστηριότητας. Έτσι για παράδειγμα, μια υπογραφή μπορεί να μετρηθεί είτε στατικά – μέσω της εξέτασης της υπογραφής μετά την αποτύπωσή της – είτε δυναμικά – μέσω της παρατήρησης της διαδικασίας της υπογραφής.

Αναλόγως με τον βαθμό που επιτρέπουν τον διαχωρισμό μεταξύ δύο ατόμων αλλά και της μονιμότητας τους τα βιομετρικά γνωρίσματα μπορούν να διακριθούν σε *soft* και *hard biometrics*. Τα soft biometrics αφορούν σε χαρακτηριστικά τα οποία παρέχουν κάποια μορφή πληροφορίας για ένα άτομο, αλλά δεν επαρκούν για τον ουσιαστικό διαχωρισμό μεταξύ δύο ατόμων. Τα χαρακτηριστικά αυτά μπορεί να είναι είτε συνεχή (όπως είναι το ύψος και το βάρος) είτε διακριτά (όπως το φύλο, το χρώμα των ματιών, η εθνικότητα). Τα hard biometrics εμπεριέχουν όλες τις ιδιότητες που απαιτούνται για τη διάκριση μεταξύ δύο ανθρώπων. Χαρακτηριστικά παραδείγματα αυτών είναι το DNA και το δαχτυλικό αποτύπωμα. Όπως είναι προφανές, η χρήση των soft biometrics για ταυτοποίηση ή επιβεβαίωση της ταυτότητας ενός ατόμου είναι εφικτή μόνο μέσω του συνδυασμού τους με hard biometrics. Σημαντική πάντως είναι η χρησιμότητα αυτών για άλλες εφαρμογές, όπως είναι η σύνδεση ενός ατόμου με κάποιο προφίλ διατηρώντας την ανωνυμία του όπως θα δούμε και σε ακόλουθη παράγραφο.

Η κύρια κατηγοριοποίηση των βιομετρικών συστημάτων η οποία απαντάται στη βιβλιογραφία είναι ο διαχωρισμός τους σε *μονοτροπικά* (*unimodal*) και *πολυτροπικά* (*multimodal*). Στην πρώτη κατηγορία εντάσσονται τα βιομετρικά συστήματα τα οποία βασίζονται σε ένα βιομετρικό χαρακτηριστικό, όπως είναι για παράδειγμα τα συστήματα αναγνώρισης προσώπου. Αντιθέτως, τα πολυτροπικά βιομετρικά συστήματα συνδυάζουν περισσότερα του ενός βιομετρικού χαρακτηριστικού με στόχο τη βελτίωση της ακρίβειας του συστήματος κατά την ταυτοποίηση, την επιβεβαίωση ταυτότητας ή τη σύνθεση του προφίλ του

ατόμου. Παραδείγματα τέτοιων συστημάτων αποτελούν συστήματα που συνδυάζουν αναγνώριση προσώπου με αναγνώριση φωνής. Τα τελευταία χρόνια απαντάται και ο όρος διατροπικά (intramodal) ο οποίος αφορά σε βιομετρικά συστήματα που εστιάζουν σε ένα βιομετρικό χαρακτηριστικό και συμπεριλαμβάνουν fusion διαφορετικών τεχνικών σε διάφορα στάδια του βιομετρικού συστήματος.



Εικόνα 2 Κατηγορίες Βιομετρικών Χαρακτηριστικών

2.6 Παραδείγματα Βιομετρικών Χαρακτηριστικών

2.6.1 Πρόσωπο

Η αναγνώριση προσώπου αποτελεί υποπεριοχή του γενικότερου προβλήματος αναγνώρισης αντικειμένων το οποίο απαιτεί διαχωρισμό μεταξύ αντικειμένων τα οποία διαφέρουν ελάχιστα μεταξύ τους (στην προκειμένη τα ανθρώπινα πρόσωπα) και για τον λόγο αυτόν συνιστά μεγάλη πρόκληση στον χώρο των προβλημάτων όρασης υπολογιστών. Πρόκειται για μια μη-επεμβατική βιομετρική μέθοδο η οποία βρίσκει εφαρμογή σε ποικίλους τομείς, όπως το εμπόριο (διασκέδαση, επεξεργασία ταινιών, βιντεοτηλέφωνα, τηλεδιάσκεψη), η ασφάλεια, η βιομηχανία και η επιβολή του νόμου.

Το πρώτο κύριο βήμα ενός συστήματος αναγνώρισης προσώπου είναι η ανίχνευση προσώπου στην εικόνα. Μια εικόνα μπορεί να περιέχει κανένα, ένα ή περισσότερα πρόσωπα. Έτσι, η διαδικασία ανίχνευσης και εντοπισμού του προσώπου στοχεύει στον αυτόματο και αξιόπιστο καθορισμό της θέσης όλων των περιοχών στην εικόνα οι οποίες περιέχουν ένα άτομο, ανεξαρτήτως της τρισδιάστατής του θέσης, του προσανατολισμού του, της θέσης του και των συνθηκών φωτός. Δεδομένου ότι η πολυπλοκότητα και η αποδοτικότητα της αναγνώρισης προσώπου εξαρτάται από την ακρίβεια της διαδικασίας εντοπισμού του στην εικόνα, η διαδικασία αυτή είναι μείζονος σημασίας. Έτσι, κάθε μέθοδος η οποία χρησιμοποιείται για τον εντοπισμό του προσώπου στην εικόνα θα πρέπει να λαμβάνει υπ’οψιν ότι απαντάται μεγάλη μεταβλητότητα όσον αφορά στο μέγεθος, στο χρώμα, στο σχήμα, στην υφή, στον προσανατολισμό και στην έκφραση του ανθρώπινου προσώπου.

Ο εντοπισμός προσώπου σε εικόνα αποτελεί ένα πρόβλημα για το οποίο έχει λάβει χώρα πυρετώδης ερευνητική δραστηριότητα καταλήγοντας σε πληθώρα μεθόδων εντοπισμού προσώπου οι οποίες μπορούν να κατηγοριοποιηθούν σε τέσσερεις κύριες κατηγορίες [26]:

- *Feature Invariant*: οι οποίες στοχεύουν στον εντοπισμό χαρακτηριστικών του προσώπου τα οποία υφίστανται ακόμη και όταν η οπτική γωνία, ο προσανατολισμός ή οι συνθήκες φωτισμού ποικίλουν. Τα ανιχνευμένα χαρακτηριστικά χρησιμοποιούνται για καθορισμό της θέσης του προσώπου. Βάση για τις μεθόδους αυτές αποτελεί η παρατήρηση ότι οι άνθρωποι έχουν τη δυνατότητα ανίχνευσης προσώπων υπό ποικίλες

συνθήκες και διαφορετικές περιπτώσεις προσανατολισμού, κι έτσι πρέπει να υπάρχουν χαρακτηριστικά ή οι ιδιότητες που δεν επηρεάζονται από αυτές τις μεταβαλλόμενες συνθήκες και καταστάσεις. Παραδείγματα τέτοιων χαρακτηριστικών αποτελούν η υφή του προσώπου (συμπεριλαμβάνοντας τις τρίχες και το δέρμα) και το χρώμα του προσώπου. Το τελευταίο χρησιμοποιείται έχοντας ως βάση μελέτες οι οποίες δείχνουν ότι η διαφορά στο χρώμα του προσώπου οφείλεται κυρίως στη διαφορά της έντασης (intensity) παρά στη διαφορά του χρώματος (chrominance) καλήγοντας σε πληθώρα μοντέλων του χρώματος του ανθρώπινου προσώπου τα οποία έχουν αναπτυχθεί σε διάφορους χρωματικούς χώρους (colour spaces). Τα τελευταία χρόνια πλήθος τεχνικών οι οποίες συνδυάζουν χαρακτηριστικά του προσώπου προκειμένουν να εντοπίσουν πρόσωπα έχουν παρουσιαστεί. Τα χαρακτηριστικά αυτά περιλαμβάνουν το χρώμα του δέρματος, το σχήμα και το μέγεθος για τον καθορισμό των υποψήφιων περιοχών στην εικόνα, ενώ ακολουθεί η φάση της επιβεβαίωσης στις υποψήφιες περιοχές κατά τις οποίες εξετάζεται η ύπαρξη χαρακτηριστικών του προσώπου όπως είναι η μύτη, το στόμα, τα φρύδια και τα μαλλία.

- *Μέθοδοι βασισμένες σε Γνώση (Knowledge-based)*: οι οποίες βασίζονται στην προσπάθεια κωδικοποίησης της ανθρώπινης γνώσης όσον αφορά στο πώς μοιάζει ένα συνηθισμένο ανθρώπινο πρόσωπο. Τις περισσότερες φορές αυτή η γνώση αποτελείται από κανόνες οι οποίοι αποτυπώνουν τις σχέσεις μεταξύ των χαρακτηριστικών του προσώπου σχετικά με τις σχετικές τους αποστάσεις και θέσεις. Παραδείγματα τέτοιων κανόνων είναι το γεγονός ότι ένα πρόσωπο έχει δύο μάτια τα οποία είναι συμμετρικά το ένα ως προς το άλλο, μια μύτη και ένα στόμα. Όμως, εμφωλεύει ο κίνδυνος αν οι κανόνες αυτοί είναι πολύ αυστηροί να μην καταφέρουν να ανιχνεύσουν ένα πρόσωπο το οποίο δεν τους ικανοποιεί, ενώ αν είναι πολύ γενικοί, μπορεί να παρέχουν πολλά λανθασμένα θετικά. Επιπροσθέτως, ο εντοπισμός προσώπων σε διαφορικές στάσεις μεγαλώνει σημαντικά το σύνολο των περιπτώσεων τις οποίες οι κανόνες θα πρέπει να ενσωματώνουν. Παρ’όλα αυτά, στις περιπτώσεις των εικόνων που περιέχουν την εμπρόσθια όψη προσώπων, οι μέθοδοι αυτές μπορούν να εντοπίσουν αρκετά αποδοτικά τα ανθρώπινα πρόσωπα.

- *Μέθοδοι βασισμένες στην εμφάνιση (Appearance-based methods)*: οι οποίες βασίζονται σε μοντέλα τα οποία προκύπτουν από ένα μεγάλο σύνολο εικόνων προσώπου ποικίλων χαρακτηριστικών που χρησιμοποιούνται για εκπαίδευση (training) με τα χαρακτηριστικά του προσώπου τα οποία προκύπτουν από τη διαδικασία μάθησης να αποθηκεύονται υπό τη μορφή είτε μοντέλων κατανομής είτε συναρτήσεων διάκρισης. Τα μοντέλα αυτά χρησιμοποιούνται για εντοπισμό προσώπου με την εφαρμογή τους στις υπό εξέταση εικόνες.

- *Μέθοδοι σύγκρισης προτύπων*: οι οποίες περιλαμβάνουν τη σύγκριση μιας εικόνας εισόδου με δεδομένα πρότυπα ενός προσώπου τα οποία περιγράφουν είτε το πρόσωπο ολόκληρο είτε τα χαρακτηριστικά του προσώπου ξεχωριστά. Κατά τη διαδικασία αυτήν λαμβάνει χώρα ο υπολογισμός των τιμών συσχέτισης και έτσι η παρουσία ενός προσώπου στην εικόνα εισόδου καθορίζεται από τις τιμές αυτές.

Η αναγνώριση προσώπου λαμβάνει υπ’όψιν το αποτέλεσμα του εντοπισμού προσώπου και στοχεύει στην ταυτοποίηση ή στην επιβεβαίωση της ταυτότητας του ατόμου που έχει εντοπιστεί στην εικόνα. Οι υπάρχουσες μέθοδοι αναγνώρισης προσώπου μπορούν να διακριθούν σε δύο μεγάλες κατηγορίες : τις αναλυτικές και τις ολιστικές μεθόδους [25] [28]. Οι αναλυτικές ή βασισμένες σε χαρακτηριστικά μέθοδοι (feature-based) εστιάζουν στη μελέτη της εξαγωγής χαρακτηριστικών στον χώρο και υπολογίζουν ένα σύνολο από γεωμετρικά χαρακτηριστικά από το πρόσωπο όπως τα μάτια, η μύτη και το στόμα. Οι ολιστικές ή βασισμένες στην εμφάνιση (appearance-based) μέθοδοι λαμβάνουν υπ’όψιν τις γενικές ιδιότητες του ανθρώπινου προσώπου. Το πρόσωπο αναγνωρίζεται συνολικά χωρίς τη χρήση μόνο συγκεκριμένων σημείων αναφοράς τα οποία λαμβάνονται από διαφορετικές περιοχές του προσώπου. Οι μέθοδοι αυτές γενικώς λειτουργούν απευθείας στην αναπαράσταση σε πίνακα της έντασης των εικονοστοιχείων (pixel intensity) χωρίς τον εντοπισμό των χαρακτηριστικών του προσώπου. Δεδομένου ότι δεν απαιτείται ανίχνευση γεωμετρικών χαρακτηριστικών του προσώπου, το σύνολο των μεθόδων αυτών είναι συνήθως πιο πρακτικό και πιο εύκολο για υλοποίηση σε σχέση με γεωμετρικές βασισμένες σε χαρακτηριστικά μεθόδους. Ερευνητικές προσπάθειες για συνδυασμό αναλυτικών και ολιστικών μεθόδων επίσης έχουν λάβει χώρα

[27].

2.6.2 Δαχτυλικά Αποτυπώματα

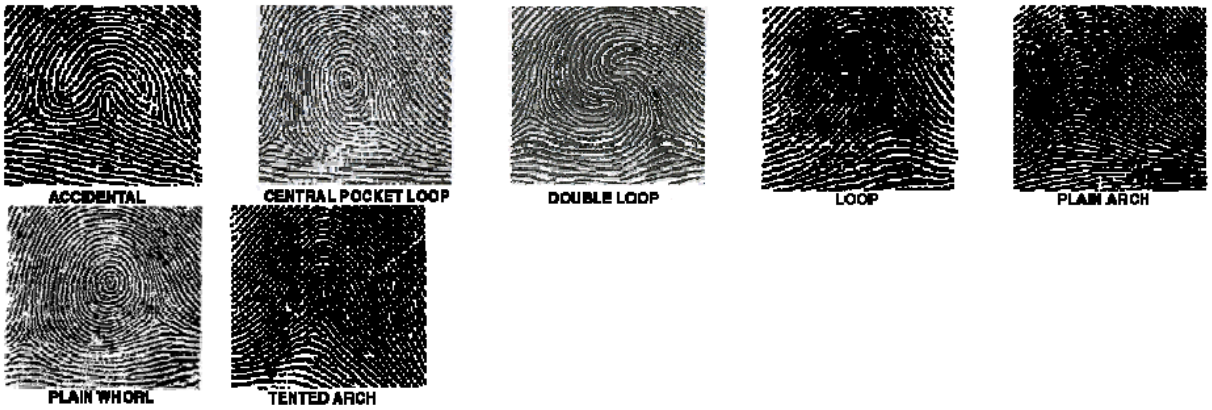
Όπως έχει ήδη αναφερθεί, η ιδέα ότι η χρήση δακτυλικών αποτυπωμάτων μπορεί να οδηγήσει σε ταυτοποίηση ή επιβεβαίωση της ταυτότητας ενός ατόμου αποτελεί μια μέθοδο η οποία απαντάται ανά τους αιώνες. Η τεχνολογική έκρηξη που παρατηρήθηκε τον 20° αιώνα έδωσε πρόσφορο έδαφος για την ανάπτυξη αυτομάτων συστημάτων αναγνώρισης δακτυλικών αποτυπωμάτων τα οποία δοκιμάστηκαν και εισήχθησαν αρχικά στις Ηνωμένες Πολιτείες Αμερικής και στην Αυστραλία. Η αρχική αυτή εισαγωγή τέτοιων βιομετρικών συστημάτων άνοιξε τον δρόμο για τη γενικότερη χρήση των δακτυλικών αποτυπωμάτων ως μέσο ταυτοποίησης και πιστοποίησης της ταυτότητας η οποία απαντάται σε πληθώρα χώρων, όπως επίσημη ταυτοποίηση ατόμων από τις αστυνομικές αρχές και ο έλεγχος πρόσβασης σε υπολογιστές μεταξύ των άλλων.

Τα δακτυλικά αποτυπώματα δεν καθορίζονται γενετικά, αλλά ξεκινούν να αναπτύσσονται τυχαία κατά τον 3° μήνα της κύησης στις άκρες των δακτύλων των εμβρύων [16]. Η τεχνική αναγνώρισης δακτυλικών αποτυπωμάτων βασίζεται σε αυτές τις ποικίλες θηλοειδείς δομές στα άκρα των ανθρώπινων δακτύλων. Η μοναδικότητα των δακτυλικών αποτυπωμάτων μεταξύ των ατόμων βασίζεται στην ποικιλία διακριτών χαρακτηριστικών, όπως είναι οι πόροι και οι ραβδώσεις. Έτσι, για παράδειγμα, το πλήθος των πόρων, τα σχετικά τους σχήματα και μεγέθη, οι σχετιές τους θέσεις αποτελούν παράγοντες που ενισχύουν τη μοναδικότητα αυτών των χαρακτηριστικών.

Ρίχνωντας μια πιο κοντινή ματιά στη μορφολογία των άκρων των δακτύλων, οι πόροι σχηματίζονται όταν οι ιδρωτοποιοί αδένες στο υποδόριο στρώμα του δέρματος παράγουν πόρους ιδρώτα οι οποίοι με τη σειρά τους μεγαλώνουν διαμέσου του υποδόριου στρώματος και του δέρματος στην επιδερμίδα [19]. Οι επιδερμικές ραβδώσεις ξεκινούν να σχηματίζονται κατά τον 6° μήνα της κύησης, οπότε και το πρόπυπο το οποίο σχηματιζόταν στην περιοχή της αδενικής πτυχής μεταφέρεται στην επιδερμίδα. Πολλοί παράγοντες επηρεάζουν τον σχηματισμό του πρότυπου της επιδερμίδας [20], όπως είναι η σταθεροποίηση η οποία λαμβάνει χώρα όταν οι πόροι έκκρισης των ιδρωτοποιών αδένων ανοίγουν στην επιφάνεια, κατά τακτά χρονικά διαστήματα, στις θηλακοειδείς ραβδώσεις. Αυτά τα ανοίγματα στην

επιφάνεια είναι πόροι και η συχνότητα εμφάνισης αυτών αποτελεί σημαντική παράμετρο που καθορίζει τη μοναδικότητα των συνθέσεων των πόρων. Η θέση των πόρων είναι σταθερή μόλις ολοκληρωθεί ο σχηματισμός τους στη ράβδωση. Έρευνες οι οποίες έχουν πραγματοποιηθεί έχουν καταλήξει ότι οι πόροι δεν εξαφανίζονται, μετακινούνται ή δημιουργούνται με το πέρασμα του χρόνου [21].

Αναλόγως το σύστημα, επτά με εννέα τυπικά πρότυπα μπορούν να διακριθούν [17], τα οποία μπορούν να κατηγοριοποιηθούν γενικώς σε τόξο (arch), έλικα (whorl) και βρόχο (loop), όπως παρουσιάζονται και στην **Error! Reference source not found.**. Πέραν των κύριων χαρακτηριστικών των δακτυλικών αποτυπωμάτων χρησιμοποιούνται και δευτερεύοντα χαρακτηριστικά όπως είναι η θέση των απολήξεων των ραβδώσεων καθώς και οι διακλαδώσεις τους.



Εικόνα 3 Επτά κοινά πρότυπα που χρησιμοποιούνται από το FBI [12]

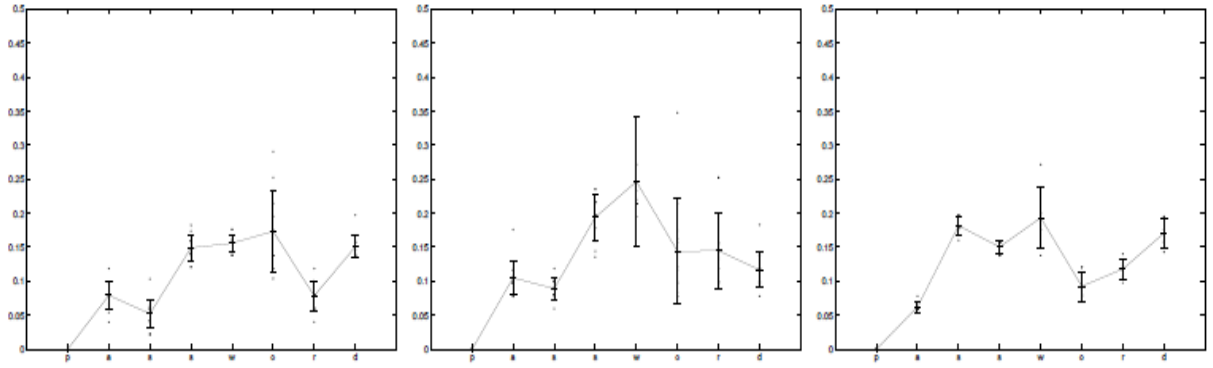
Η λήψη των δακτυλικών αποτυπωμάτων μπορεί να γίνει με διαφόρων ειδών αισθητήρες, όπως όπως οπτικούς αισθητήρες, αισθητήρες ηλεκτρομαγνητικού πεδίου, θερμικούς αισθητήρες και υπερηχητικούς αισθητήρες. Όπως θα δούμε και στην **Error! Reference source not found.**, νέοι εξελιγμένοι αισθητήρες οι οποίοι ανιχνεύουν κατά τη διάρκεια της λήψης των δακτυλικών αποτυπωμάτων αν τα δεδομένα λαμβάνονται από ζώντα οργανισμό (liveness detection) έχουν αναπτυχθεί με στόχο την αποφυγή περιπτώσεων «εξαπάτησης» των

συστημάτων αναγνώρισης δακτυλικών αποτυπωμάτων με την παρουσίαση ψεύτικων ή κλεμμένων δακτυλικών αποτυπωμάτων.

2.6.3 Δυναμικές της Πληκτρολόγισης

Τις τελευταίες δύο δεκαετίες, οι υπολογιστές έχουν ένα πολύ σημαντικό μέρος της ζωής μας και της κοινωνίας μας. Όσο περισσότερο τα χρησιμοποιούν, τόσο πιο δύσκολο γίνεται να αγνοήσουμε την ανάγκη της αξιοπιστίας και της ασφάλειας περαιτέρω σχετικά με τις εφαρμογές και τις υπηρεσίες που χρησιμοποιούμε. Κάθε μέρα, εμείς χρησιμοποιούμε τους υπολογιστές να έχουν πρόσβαση e-mail μας, περιηγηθείτε φόρουμ, παίζουν video games, αλλά και να διαχειρίζονται τους τραπεζικούς λογαριασμούς μας και άλλα ευαίσθητα δεδομένα. Έτσι, καταλαβαίνουμε ότι πρόκειται για παραβίαση λογαριασμού μπορεί να προκαλέσει πολλά προβλήματα, και όχι μόνο στον ιδιοκτήτη του λογαριασμού. Αυτό είναι όπου η δυναμική πληκτρολόγισης μπαίνει βιομετρικών

Τα τελευταία χρόνια, έρευνες έχουν δείξει ότι κάθε άτομο έχει ορισμένα μοναδικά χαρακτηριστικά, τα οποία υπολογίζονται από τον ρυθμό πληκτρολογώντας το πληκτρολόγιο του. Η καθυστέρηση μεταξύ κάθε πάτημα πλήκτρου ή τη διάρκεια που κατέχει ένα συγκεκριμένο πατημένο το πλήκτρο κατά την πληκτρολόγηση μπορεί να τον διακρίνει μεταξύ άλλων χρηστών. Είναι ένα βιομετρικό συμπεριφοράς που μπορεί να χρησιμοποιηθεί με πολλούς τρόπους, όπως για την ανίχνευση συναισθήματα, αλλά το σημαντικότερο, για την επικύρωση ή την ταυτοποίηση ακόμη και ένα άτομο.



Εικόνα 4 Χρόνοι αναμονής μεταξύ πληκτρολογήσεις όταν γράφετε τη λέξη "password" από τρεις διαφορετικούς ανθρώπους. **[20]**

Για παράδειγμα, μπορεί να χρησιμοποιηθεί σε μια διαδικασία ελέγχου ταυτότητας κωδικού πρόσβασης. Το σύστημα κωδικό, είναι ένα από τα πιο ευρέως χρησιμοποιούμενα τρόπους για τον έλεγχο ταυτότητας ενός προσώπου, ειδικά στο διαδίκτυο. Αυτός είναι ο λόγος πληκτρολόγησης ευαίσθητα authentications κωδικό μπορεί να αυξήσει την ασφάλεια του συστήματος δραματικά. Φανταστείτε ένα σύστημα ελέγχου ταυτότητας που δεν απαιτούν μόνο το σωστό κωδικό πρόσβασης, αλλά και τη σωστή ρυθμό δακτυλογράφησης να δεχθεί το χρήστη. Λειτουργεί με τον ίδιο τρόπο όπως η πιστοποίηση υπογραφής σε χειρογράφου βιομετρικά χαρακτηριστικά.

2.7 Βιομετρικές εφαρμογές

Οι βιομετρικές εφαρμογές μπορούν γενικώς να κατηγοριοποιηθούν σε:

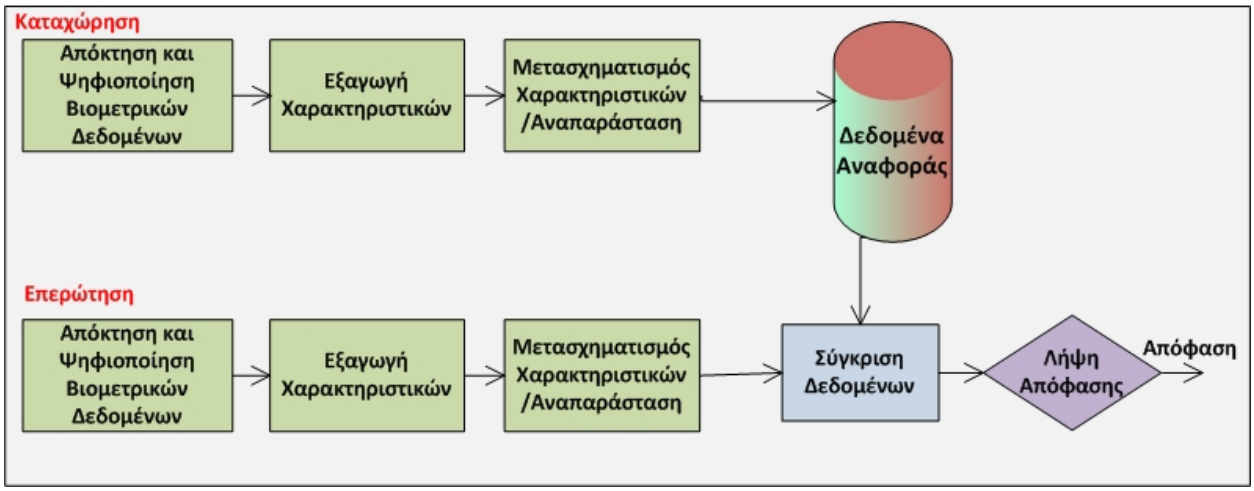
- ***Κυβερνητικού ελέγχου μοντέλα ταυτοποίησης***: για τις εφαρμογές αυτές μια δημόσια αρχή αναλαμβάνει να συλλέξει βιομετρικά δεδομένα προκειμένου να τα συμπεριλάβει σε μια εφαρμογή ταυτοποίησης, όπως είναι οι ταυτότητες, τα διαβατήρια και οι κάρτες κοινωνικής ασφάλισης. Ο έλεγχος επί των δεδομένων αυτών μπορεί να είναι κεντρικός, να διαιρείται σε περισσότερους του ενός οργανισμούς αλλά υπό κατάλληλες συμφωνίες ή πολυμερής χωρίς τις απαραίτητες συμφωνίες σχετικά με την αποκάλυψη ή τη μεταφορά των βιομετρικών δεδομένων.
- ***Μοντέλο ελέγχου πρόσβασης***: σε αυτήν την κατηγορία μια δημόσια ή ιδιωτική αρχή αναλαμβάνει τη συλλογή των βιομετρικών δεδομένων προκειμένου να διασφαλίσει την

πρόσβαση σε μια τοποθεσία ή σε μια διαδικτυακή εφαρμογή. Ο έλεγχος επί των δεδομένων μπορεί να είναι κεντρικός ή διαιρεμένος σε περισσότερους του ενός οργανισμούς αλλά με κατάλληλες συμφωνίες ή να διαιρείται έτσι ώστε το άτομο να μοιράζεται τον έλεγχο.

- ***Μικτό μοντέλο*** : στο μοντέλο αυτό τα βιομετρικά δεδομένα που έχουν συλλεγεί θα μοιράζονται και να ανταλλάσσονται μετακύ δημόσιων και ιδιωτικών αρχών.
- ***Μοντέλο διευκόλυνσης*** : είτε το υποκείμενο αναλαμβάνει εξ ολοκλήρου την απόφαση χρήσης των βιομετρικών δεδομένων για αποκλειστικά ιδιωτικούς σκοπούς διευκόλυνσης, είτε ένας οργανισμός χρησιμοποιεί τα βιομετρικά δεδομένα προς απλούστευση της διαχειριστικής διαδικασίας είτε με κεντρικό είτε με διαμοιρασμένο έλεγχο.
- ***Μοντέλο Παρακολούθησης*** : μια δημόσια ή ιδιωτική αρχή αναλαμβάνει να συλλέξει και να επεξεργαστεί βιομετρικά δεδομένα για σκοπούς παρακολούθησης.

2.8 Οι δύο φάσεις λειτουργίας των Βιομετρικών Συστημάτων

Όλα τα βιομετρικά συστήματα λειτουργούν σε δύο ξεχωριστές φάσεις; την *καταχώρηση* των βιομετρικών δεδομένων αναφοράς (*enrolment*) και την *επερώτηση* του βιομετρικού συστήματος (**Error! Reference source not found.**). Η πρώτη φάση, αποτελεί μέρος της αρχικοποίησης του βιομετρικού συστήματος όταν πια αυτό έχει τεθεί σε λειτουργία. Κατά τη φάση αυτή παρέχονται στο σύστημα δείγματα των βιομετρικών χαρακτηριστικών των ατόμων προκειμένου να παραχθούν τα βιομετρικά δεδομένα αναφοράς (reference template) μετά την αποθήκευση των οποίων το άτομο θεωρείται γνωστό στο βιομετρικό σύστημα.



Εικόνα 5 Γενικά Στάδια Βιομετρικού Συστήματος

2.9 Είδη Σφαλμάτων Βιομετρικών Συστημάτω

Τα συστηματικά και στατιστικά σφάλματα των μετρήσεων και των αλγόριθμων των διαδικασιών εξαγωγής των βιομετρικών χαρακτηριστικών και σύγκρισης καθορίζουν τα όρια της εφαρμογής ενός βιομετρικού συστήματος και την ικανότητα διαχωρισμού μεταξύ διαφορετικών ατόμων. Μια σειρά από μέτρα τα οποία έχουν οριστεί προκειμένου να αποτιμήσουν την απόδοση των βιομετρικών συστημάτων. Τα μέτρα αυτά αφορούν σε είδη σφαλμάτων τα οποία μπορεί να λάβουν χώρα όσον αφορά στην εκτίμηση του βιομετρικού συστήματος.

- **False Acceptance Rate (FAR)**

  Το FAR αντιπροσωπεύει την πιθανότητα ότι ένα σύστημα ελέγχου ταυτότητας μπορεί να παρέχει πρόσβαση σε ένα απατεώνα. Η πιθανότητα να συμβεί κάτι τέτοιο σε ένα σύστημα ελέγχου ταυτότητας που λειτουργεί με φυσικό βιομετρικών στοιχείων είναι πολύ χαμηλές, αλλά στην περίπτωση των βιομετρικών στοιχείων συμπεριφοράς, δεν είναι κάτι ασυνήθιστο, ανάλογα με την αυστηρότητα του συστήματος. Ειδικά αν ο απατεώνας είχε την ευκαιρία να παρακολουθήσουν και να μάθουν πώς να αντιγράψετε τον τρόπο του αρχικού χρήστη της ταυτότητας.

- **False Rejection Rate (FRR)**

Το FRR αντιπροσωπεύει την πιθανότητα ενός συστήματος ελέγχου ταυτότητας αρνείται την πρόσβαση στο αρχικό χρήστη. Πάλι, η πιθανότητα να συμβεί αυτό σε ένα σύστημα που χρησιμοποιεί συμπεριφορικές βιομετρικών δεδομένων είναι πολύ υψηλότερη από ό, τι με τη σωματική βιομετρία. Ειδικά αν ο χρήστης δεν είναι σε κανονική κατάσταση του νου, όπως είναι κουρασμένοι του, σε συναισθηματικές διαταραχές, ή υπό την επήρεια αλκοόλ, κλπ., Επίσης, εξαρτάται από την αυστηρότητα του συστήματος.

- **False Match Rate (FMR)**

  Το FMR αντιπροσωπεύει την πιθανότητα ενός συστήματος αναγνώρισης με ψευδώς θετικά του αγώνα από ένα συγκεκριμένο πρόσωπο. Η πιθανότητα να συμβεί αυτό εξαρτάται κυρίως από τους αλγόριθμους που χρησιμοποιούνται, καθώς και την πολυπλοκότητα των βιομετρικών χαρακτηριστικών που χρησιμοποιούνται. Για παράδειγμα, η αναγνώριση προσώπου και ταυτοποίηση είναι μία από τις πιο περίπλοκες διαδικασίες, ιδίως όταν δεν ελέγχονται οι συνθήκες τοποθέτησης, τον προσανατολισμό, την τοποθεσία και το φως.

- **False Non-Match Rate (FNMR)**

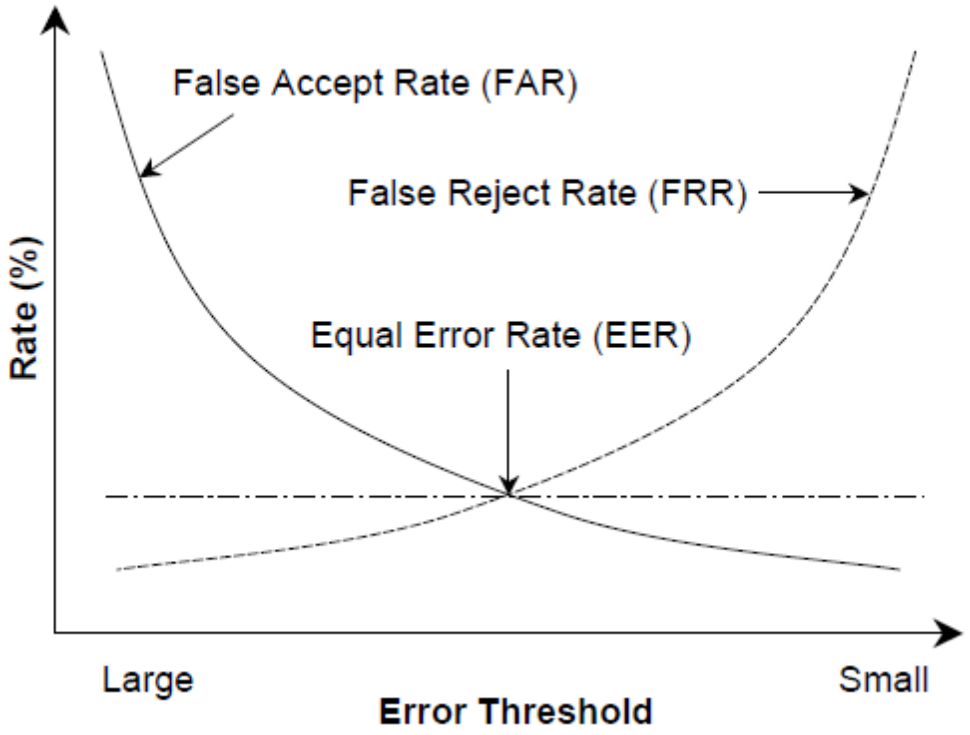
  Το FNMR αντιπροσωπεύει την πιθανότητα ένα σύστημα αναγνώρισης δεν είναι σε θέση να προσδιορίσει σωστά το άτομο, ενώ έχουν βιομετρικών δεδομένων του. Ακριβώς όπως και το ποσοστό εσφαλμένης αγώνα, η πιθανότητα να συμβεί αυτό εξαρτάται από τους αλγόριθμους που χρησιμοποιούνται, η πολυπλοκότητα των βιομετρικών χαρακτηριστικών, καθώς και τις εξωτερικές συνθήκες, όταν και τα δύο στοιχεία συνελήφθησαν.

- **Equal Error Rate (EER)**

  Το EER μπορεί να αλλάξει από τα στατιστικά ποσοστά σφάλματος και αλγόριθμοι που χρησιμοποιούνται, ανάλογα με το τι θέλουμε να επιτύχουμε, και ό, τι θεωρούμε σημαντικό σε μια διαδικασία πιστοποίησης ή αναγνώρισης. Κατά τη διαδικασία ελέγχου ταυτότητας, θα μπορούσαμε, για παράδειγμα, δεν νοιάζονται για το ποσοστό εσφαλμένης απόρριψης, γιατί θέλουμε το σύστημά μας να είναι εξαιρετικά ασφαλής. Ή ίσως να μην είναι πολύ ενδιαφέρονται για τη διατήρηση του συστήματος είναι πολύ ασφαλές, και να μειώσει την αυστηρότητα στην οποία το σύστημα δέχεται ένα χρήστη.

Το ίδιο ισχύει και για τη διαδικασία ταυτοποίησης, το FMR και το FNMR. Η πιθανότητα κατά την οποία και οι δύο τύποι σφαλμάτων FAR, FRR, και FMR, FNMR εκ τούτου, έχουν ίσα ποσοστά, καλείται η ίση ποσοστό σφάλματος. Όταν για παράδειγμα, τόσο η FAR και ο FRR είναι 3%.

Ένα τελευταίο σημαντικό πράγμα που πρέπει να ξέρουμε είναι ότι FAR, FRR και FMR, ανάλογα FNMR είναι αντίστροφη ιδιότητες. Σε ένα σύστημα ελέγχου ταυτότητας για παράδειγμα, η ασφαλέστερα είναι, ως εκ τούτου, τόσο χαμηλότερη είναι η FAR είναι, τόσο υψηλότερη είναι η FRR θα είναι επίσης, και το αντίθετο.



Εικόνα 6 Σχέσεις μεταξύ FAR FRR EER [21]

Έννοιες

Στον πίνακα που ακολουθεί, οι όροι που χρησιμοποιήθηκαν στην διατριβή και το πείραμα εξηγούνται:

API	Application Programming Interface
BEM	Biometric Evaluation Methodology
BIR	Biometric Identification Record
BLR	Biometric Live Record
BMG	Biometrics Grid
CER	Crossover Error Rate
DNA	Deoxyribonucleic acid
EER	Equal Error Rate
EM	Expectation Maximization
EPIC	Electronic Privacy Information Center
EPSRC	Engineering and Physical Sciences Research Council
FAR	False Acceptance Rate
FER	Failure to Enroll Rate
FERET	Facial Recognition Technology
FIDIS	Future of Identity in the Information Society
FMR	False Match Rate
FN	False Negative
FNMR	False Non-Match Rate
FRR	False Rejection Rate
FP	False Positive
GAR	Genuine Accept Rate

GMM	Gaussian Mixture Model
HMM	Hidden Markov Models
ISECOM	Institute for Security and Open Methodologies
ML	Maximum Likelihood
MRTDs	Machine Readable Travel Documents
OAE	Otoacoustic Emission
QoS	Quality of Service
ROC	Receiver Operating Characteristic
SLA	Service Level Agreement
SVM	Support Vector Machines
TN	True Negative
TP	True Positive
URL	Unified Resource Locator
VNTR	Variable Number of Tandem Repeats
XM2VTSDB	Extended Multimodal *Verification for Teleservices and Security* applications Database

Βιβλιογραφία

[1] Jain, A. K.; Ross, Arun; Prabhakar, Salil (January 2004), "An introduction to biometric recognition", IEEE Transactions on Circuits and Systems for Video Technology 14th (1): 4–20.

[2] Bromme, A., A classification of biometric signatures, Multimedia and Expo, 2003, ICME '03, Proceedings. 2003 International Conference on Publication Date: 6-9 July 2003 volume: 3, page(s): III- 17-20 vol.3.

[3] The Biometric Consortium, Introduction to Biometrics, available at: http://www.biometrics.org/introduction.php, accessed on 29th April 2009.

[4] EPSRC, Otoacoustic Emission Based Biometric Systems, available at: http://gow.epsrc.ac.uk/ViewGrant.aspx?GrantRef=EP/E015522/1, accessed on 29th April 2009.

[5] Janeen Renaghan, “Etched in Stone,” Zoogoer, August 1997, (Smithsonian National Zoological Park, 26 January 2005).

[6] “Dermatoglyphics,” Hand Analysis, International Institute of Hand Analysis, 24 January 2005.

[7] Z. McMahon, Biometrics: History, Indiana University, Indiana University Computer Science Department, 24 January 2005, available at http://www.cs.indiana.edu/~zmcmahon/biometrics-history.htm

[8] Alphonse Bertillon, access on 4th May 2009, available at: http://en.wikipedia.org/wiki/Bertillon

[9] Bonsor, K.. "How Facial Recognition Systems Work", accessed on 4th May 2009, available at http://electronics.howstuffworks.com/facial-recognition.htm

[10] von Hardenberg, I., "Warum Neugeborene mehr wissen, als Große manchmal ahnen", GEO (7), pp.27-42, Hamburg, 2001.

[11] Amberger, M., Fischer, S., Rößler, J., "Biometrische Verfahren – Studie zu State of the Art", Erlangen, 2003, available at: http://www.wi3.unierlangen.de/forschung/Biometrie_StateOfTheArt/Biometrie.pdf

[12] Mark Gasson, Martin Meints, Kevin Warwick, "D3.2: A study on PKI and biometrics", 2005, available at : http://www.fidis.net/resources/deliverables/hightechid/#c1785

[13] Webster, J.G., ed., Medical Instrumentation, 2nd ed. Boston: Houghton Mifflin Company, 1992, Chapter 5, pp. 249-250.

[14] Hirsch, W. and J.U. Schweichel, "Morphological Evidence Concerning the Problem of Skin Ridge Formation," Journal of Mental Deficiency Research, Vol. 17, pp. 58-72, 1973.

[15] Locard, E., "Les Pores et L'Identification des Criminels," Biologica, revue scientifique de medecin, Vol. 22, pp. 357-362, 1912.

[16] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 15, No. 10, pp.1042-1052, 1993.

[17] Ming-Hsuan Yang, David J. Kriegman, Narendra Ahuja, "Detecting Faces in Images: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 34-58, Jan. 2002.

[18] K. M. Lam and H. Yan, “An Analytic-to-Holistic Approach for Face Recognition based on a Single Frontal View,” IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 20, No. 7, pp.673-686, 1998.

[19] Tolba, El-Baz, El-Harby, “Face Recognition – A Literature Review”, International Journal of Signal Processing, Vol.2, No.2, 2005, pp.88-103.

[20] J.Ilonen, “Keystroke dynamics” Lappeenranta University of Technology, Skinnarilankatu 34, 53850 Lappeenranta, Finland.

[21] N. Clarke, S. Furnell, and P. Reynolds, “Biometric Authentication for Mobile Devices,” in Proceedings of the 3rd Australian Information Warfare and Security Conference 2002, 2002, pp. 61