Ministry of Higher Education and Scientific Research
University of Tunis El Manar
National Engineering School of Tunis
Academic Year: 2018-2019

THESIS

FOR THE AWARD OF THE DEGREE OF :

DOCTOR OF PHILOSOPHY OF UNIVERSITY OF TUNIS EL MANAR

DELIVERED BY :

NATIONAL ENGINEERING SCHOOL OF TUNIS

DISCIPLINE :

INFORMATION AND COMMUNICATION SCIENCES AND TECHNOLOGIES

PRESENTED AND DEFENDED BY :

ABIR MHENNI

NATIONAL DIPLOMA OF ENGINEER IN APPLIED COMPUTER SCIENCE, ENISO
MASTER OF RESEARCH DEGREE IN INTELLIGENT AND COMMUNICATING SYSTEMS, ENISO

ON :04 APRIL 2019

CONTRIBUTION TO THE BIOMETRIC TEMPLATE UPDATE: APPLICATION TO KEYSTROKE DYNAMICS MODALITY

Committee

CHAIR:	MRS. HENDA HAJJAMI BEN GHEZALA	Professor, Mannouba University
REVIEWER:	MR. HASSEN SEDDIK	Professor, Tunis University
REVIEWER:	MR. PATRICK BOURS	Professor, Norwegian University
EXAMINER:	MRS. AFEF KACEM	Professor, Tunis University
DIRECTOR:	MRS. NAJOUA ESSOUKRI BEN AMARA	Professor, University of Sousse
CO-DIRECTOR:	MR. CHRISTOPHE ROSENBERGER	Professor, University of Caen
CO-SUPERVISOR:	MRS. ESTELLE CHERRIER	Professor, University of Caen

---

DOCTORAL SCHOOL :

ENGINEERING SCIENCE & TECHNOLOGY

RESEARCH LABORATORY :

LATIS- LABORATORY OF ADVANCED TECHNOLOGY AND INTELLIGENT SYSTEMS

GREYC- GROUPE DE RECHERCHE EN INFORMATIQUE, IMAGE, AUTOMATIQUE ET INSTRUMENTATION DE CAEN

This thesis is dedicated to:

My great parents Abdelmajid & Aimia,

who never stop giving of themselves in countless ways;

My dearest brothers Alaa & Azmi,

who stand by me with light of hope and love;

My beloved grandparents, my aunts, my uncles;

My beloved cousins, my first friends, and my whole family;

My very special friends, my dearest sisters;

All my friends and everyone I love;

May this thesis bring you appreciation and gratitude.

Abir

Acknowledgements

I wish to express my sincere appreciation and gratitude to those who have contributed to this thesis and supported me in one way or the other.

First of all, I am extremely grateful to my main thesis director Mrs Najoua ESSOUKRI BEN AMARA, Professor at National Engineering School of Sousse. Despite her busy schedule, she would always had the time to discuss my thesis progress and to reply to my e-mails, even at late hours. I especially would like to thank her for her patience and support in overcoming numerous obstacles I have been facing through my research. Her guidance led me to accept nothing less than excellence.

Very special thanks also to my co-director, Mr. Christophe Rosenberger, Professor at National Superior School of Engineering of Caen. Regardless of the distance, he always had time for me, where we would have discussion and exchange of ideas. His professional and high quality supervision had contributed to my achievement. I therefore warmly wish to express my highest appreciation of his style of supervision and I hope that I will one day emulate him.

I am also indebted to my advisor, Mrs. Estelle Cherrier, Professor at National Superior School of Engineering of Caen. She had played a very significant role to my success with her encouragements and her kind advice. Indeed, she always supported me when I needed help.

I am furthermore grateful to my thesis committee for accepting to judge my work.

I thank Mrs Henda Hajjami Ben Ghezala, Professor at National School of Computer Sciences, for agreeing to chair my thesis committee.

I would like to warmly thank Mr. Hassen Seddik, Professor at National Higher School of Engineers of Tunis, and Mr. Patrick Bours, Professor at Faculty of Information Technology and Electrical Engineering of Norway, for agreeing to evaluate my thesis work. May they find here the expression of my deep gratitude.

My sincere thanks to Mrs Afef Kacem, Professor at High School of Sciences and Techniques of Tunis, to have agreed to consider this work.

Many thanks must also go to my colleagues, especially those member of the LATIS laboratory: "Ines, Ibtissem, Imen, Fayza, Kalthoum, Lamia, Sameh, Oussama, Souleimen, Samira, Intissar, Syrine, Hiba, Jihen, Khaoula, Yosra, Imen, Nesrine, Jaghjouma, Fatma, Olfa" with whom I spent very special moments that will be etched in the memories.

I also thank my dear "Najla", administrative manager of the LATIS Laboratory for her unequaled support.

Although, I had the opportunity to work in GREYC laboratory for several weeks on two occasions, even for a short time span, the achieved work was fruitful. Thanks to all staff members of the GREYC who had treated me as if I was one of them, I very much appreciate it.

I would like to take this opportunity to say warm thanks to all my beloved friends, who have been so supportive along the way of doing my thesis.

Last but not least, deepest thanks go to all people who took part in making this thesis real.

Contents

List of Figures

List of Tables

List of Algorithms

List of Publications

The contributions of this thesis have been the subject of the following scientific publications:

Journal Papers

A. Mhenni, E. Cherrier, C. Rosenberger, and N. Essoukri Ben Amara, "Double serial adaptation mechanism for keystroke dynamics authentication based on a single password", Computers & security, accepted on 04/02/2019.

A. Mhenni, E. Cherrier, C. Rosenberger, and N. Essoukri Ben Amara, "Analysis of Doddington zoo classification for user dependent template update: Application to keystroke dynamics recognition", Future Generation Computer Systems, accepted on 20/02/2019.

P. H. Pisani, A. Mhenni, R. Giot, E. Cherrier, N, Poh, A. Carvalho, C. Rosenberger, and N. Essoukri Ben Amara, "Adaptive Biometric Systems: Review and Perspectives," ACM Computing Surveys, submitted on 12/09/2017, revised on 25/01/2019.

Conference Papers

A. Mhenni, E. Cherrier, C. Rosenberger, and N. Essoukri Ben Amara, "User Dependent Template Update for Keystroke Dynamics Recognition," in 17th Cyberworlds Conference (CW), 2018 (***Rank B, received Best Full Paper Award***).

A. Mhenni, E. Cherrier, C. Rosenberger, and N. Essoukri Ben Amara, "Adaptive Biometric Strategy using Doddington Zoo Classification of User's Keystroke Dynamics," in 14th International Wireless Communications and Mobile Computing Conference (IWCMC), 2018 (***Rank B***).

A. Mhenni, E. Cherrier, C. Rosenberger, and N. Essoukri Ben Amara, "Towards a secured authentication based on an online double serial adaptive mechanism of users' keystroke

dynamics," in 12th International Conference on Digital Society and eGovernments (ICDS), 2018 (***Rank C***).

A. Mhenni, E. Cherrier, C. Rosenberger, and N. Essoukri Ben Amara, "Keystroke template update with adapted thresholds," in Advanced Technologies for Signal and Image Processing (ATSIP), in 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), 2016.

Abbreviations

ANSI: National Agency for Computer Security

AUC: Area Under Curve

DET: Detection Error Tradeof

EER: Equal Error Rate

FAR: False Acceptance Rate

FIFO: First In First Out

FMR: False Match Rate

FNMR: False Non-Match Rate

FRR: False Rejection Rate

HTER: Half Total Error

LFU: Least frequently used

LRU: Least recently used

ROC: Receiver Operating Characteristic Curve

RR : Recognition Rate

RC: Robustness Curve

SVM: Support Vector Machine

NN: Neural Network

KNN: K Nearest Neighbor

GA: Genetic Algorithm

ANIA: Average Number of Impostor Actions

ANGA: Average Number of Genuine Actions

Abstract

The security of password based applications is one of major concerns nowadays regarding the proliferation of web and mobile applications. Keystroke dynamics is an emerging solution to reinforce logic authentication. It is a behavioral modality that verifies the typing manner of the user in addition to the verification of the syntactic conformity of the password.

The major disadvantage of the keystroke dynamics modality is the instability of the typing manner of users over time. In fact, it changes according to different causes like the emotional state, activeness, password mastery etc. Adaptive strategies are one of the most interesting solutions to handle these problems. They consist in updating the reference describing the typing rhythm of the user during the use of the authentication system.

Different contributions are proposed in this PhD thesis. First, we put forward an original adaptation strategy for keystroke dynamics that requires a single sample during enrollment. This solution considerably improves the usability of keystroke dynamics modality, since all methods in the state of the art require a lot of samples during the enrollment step. Second, throughout the operational use of the authentication system, the user's template in addition to acceptance and adaptation decision thresholds are updated; whereas in the literature, only the user's template is generally updated when keystroke dynamics modality is dealt with. Thus, we propose a novel adaptation decision called "the adapted thresholds". This new solution, compared to other approaches in the literature, leads to better results on significant datasets used by the scientific community in the field. Afterwards, an adaptive strategy for each category of users is proposed. In fact, a user dependent adaptative strategy based on Doddington Zoo theory has demonstrated competitive performances.

Keywords: Biometrics, keystroke dynamics, adaptive strategies, authentication; password security; Doddington Zoo; users classification.

Résumé

La sécurisation des applications par mot de passe est l'une des préoccupations majeures de nos jours vu la prolifération des applications web et mobiles. La dynamique de frappe au clavier est une solution émergente pour renforcer l'authentification logique. C'est une modalité biométrique comportementale qui vérifie la dynamique de frappe au clavier de l'utilisateur en plus de la vérification de la conformité syntaxique du mot de passe.

Le principal inconvénient de la modalité de la dynamique des frappe au clavier est l'instabilité de la façon de taper au clavier des utilisateurs au fil du temps. En fait, elle change en fonction de différentes causes comme l'état émotionnel, l'activité, la maîtrise des mots de passe, etc. Les stratégies adaptatives sont l'une des solutions les plus intéressantes pour remédier à ces problèmes. Elles consistent à mettre à jour la référence décrivant la dynamique de frappe de l'utilisateur au cours de l'utilisation du système d'authentification.

Différentes contributions sont proposées dans cette thèse. Premièrement, nous proposons une stratégie d'adaptation originale pour la dynamique de frappe au clavier qui nécessite un seul échantillon lors de l'enrôlement. Cette solution améliore considérablement la facilité d'utilisation de la modalité, car toutes les méthodes de l'état de l'art nécessitent beaucoup d'échantillons dans l'étape d'enrôlement. Deuxièmement, durant l'authentification, en plus du modèle de l'utilisateur, les seuils d'acceptation et d'adaptation sont mis à jour; peu de méthodes de l'état de l'art font ce type de mise à jour. Nous proposons un nouveau critère de décision de mise à jour appelé "les seuils adaptés". Cette nouvelle solution, comparée à d'autres approches de la littérature, conduit à de meilleurs résultats sur des ensembles de données significatifs utilisés par la communauté scientifique. Ensuite, une stratégie adaptative pour chaque catégorie d'utilisateurs est proposée. En fait, une stratégie d'adaptation basée sur la théorie du zoo de Doddington a démontré des performances compétitives.

Mots-clés: biométrie, dynamique de frappe au clavier, stratégies adaptatives, authentification; sécurité par mot de passe; Zoo de Doddington; classification des utilisateurs.

CHAPTER I

Introduction

Secret passwords are the most common mechanism for authenticating human users of computer systems, especially on the Internet. Passwords are the foundation of the security policy of a wide range of online services, protecting financial transactions, health records and personal communications, and blocking intrusions into corporate, electrical and military networks. Password security has become an important topic of research because of the pervasiveness of modern Web services and their increasingly critical nature.

Despite numerous research on alternative authentication schemes, text passwords have several advantages: common usage, easy implementation, no additional sensors. Thus, we cannot imagine application security without passwords. Two-factor user authentication, which is generally based on password access associated with password managers or biometrics, promises to increase the security and safety of passwords against attacks.

I.1 Motivations

The huge amount of personal and / or confidential information stored in our electronic devices, both PCs and smartphones, requires to protect them against unauthorized access. Since virtual environments are responsible for storing sensitive information and performing critical actions, security services that aim to neutralize threats on identity theft have been developed and must now evolve faster than the methods that seek to address them.

User authentication is a security service used to counter the impersonation (that is, someone who claims to be someone else) and its purpose is to protect the system against all unauthorized accesses. There are three authentication approaches that differ in how the user can prove that he is what he claims to be: proof by possession, by knowledge, or by property. The possession factor is something the user has, like a smart card; the knowledge factor of authentication is something the user knows, such as a username and password; and the property factor is a human characteristic (biometrics).

Access control methods based on knowledge and possession are the most widespread despite their weaknesses. Indeed, passwords can be forgotten, heard or guessed using various methods such as dictionary or brute force attacks. Smart cards also can be lost, stolen or cloned [Bergadano et al., 2003]. Due to these limitations, biometric authentication methods offer an alternative layer of security or added to the methods mentioned above to rely on traits inherent to the person and therefore can not be removed or easily imitated, all by remaining intuitive to the user and keeping the process convenient and efficient [Alsultan & Warwick, 2013].

In what follows, we present statistics of the various hacking attacks reported in Tunisia and around the world, as well as the possible solutions to overcome them.

I.1.1 Vulnerability of passwords to attacks

Today, in Tunisia and around the world, the official websites of governments and companies as well as accounts of social networks, e-commerce sites and e-banking have become the target of hackers. According to the latest news in relation to activism movements, the Tunisian National Agency for Computer Security (ANSI) has just raised the national alert level to 3.

According to data collected by ANSI, an average of 226,000 cybernetic attack events per month were recorded in Tunisia during the year 2016. These attacks continue to evolve and worsen in 2017 as shown in the figure I.1.

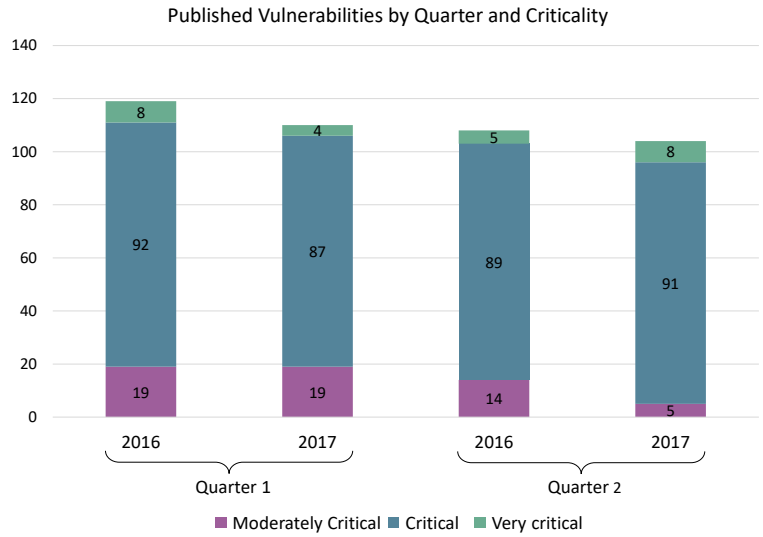


Figure I.1 – Comparison of attacks number detected by ANSI during 2016 and 2017 in Tunisia

According to "HACKMAGEDDON Information Security Timelines and Statistics" [Passeri, 2017] which publishes aggregated statistics related to different attacks events, a total of 1061 cyber-attacks were collected in 2016. During the year 2017, 950 attack events were gathered. Collected data for the statistics are derived from timelines published every two weeks (plus or minus). The journal collects major cyber events from related months selected from open sources (such as blogs or news sites) which gives an idea of the threat landscape and the main trends around the world. Of course, each event reports the sources for the sake of completeness.

These attacks usually aim at modifying a part or the entirety of a website or putting it completely out of order. In addition, they are interested in stealing information, such as industrial secrets, intellectual property or bank data (such as credit cards). The breakdown of

the number of incidents by type of attack reported in 2017 by ANSI is shown in figure I.2 and those collected by HACKMAGEDDON are illustrated in figure I.3.

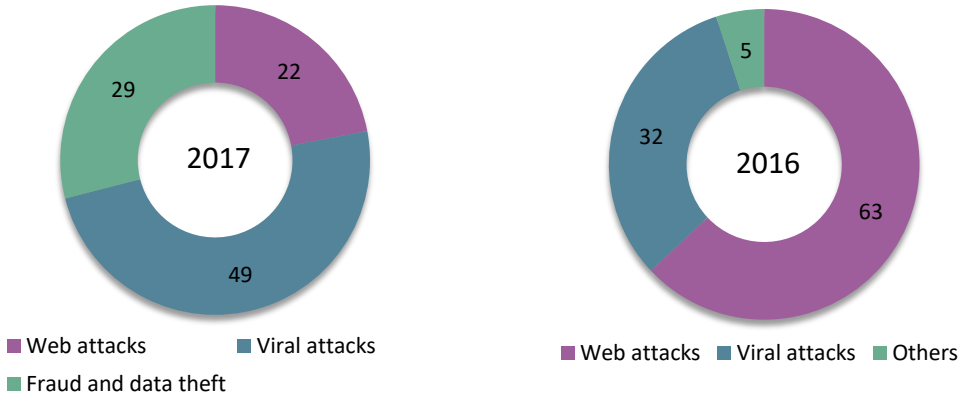


Figure I.2 – Distribution of the main incidents by type of attack according to ANSI

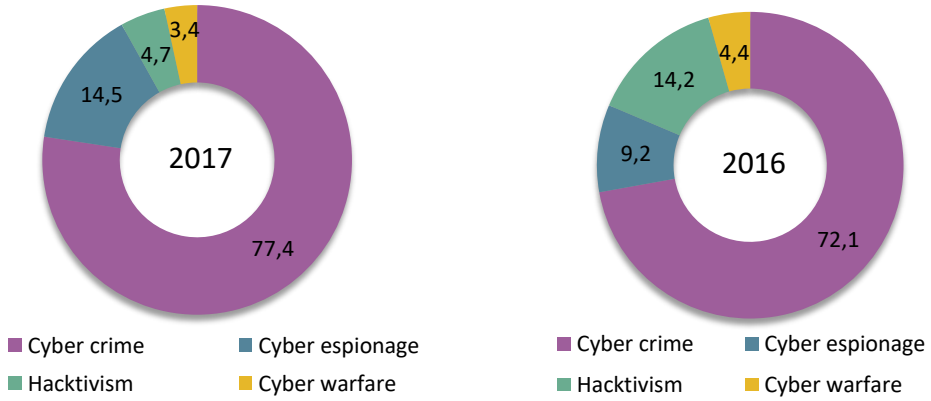


Figure I.3 – Distribution of the main incidents by type of attack according to HACKMAGEDDON

It is also interesting to compare motivations in 2016 to those of 2017. Regarding HACKMAGEDDON statistics, Cyber Crime confirmed its crown even in 2017 with a similar percentage (77.4% vs. 72.1%). Hacktivism fell to 4.3% in 2017, compared with 14.2% in 2016. In contrast, cyber espionage experienced an opposite trend, going from 4.3% to 14.5%. Cyber warfare reported a slight decrease to 3.4% of 4.3%. Concerning ANSI statistics, 2016 was the year of web attacks in Tunisia. Whereas in 2017, the percentages of viral attacks is the highest.

During 2017, 950 events were collected in comparison to 1061 events in 2016. Despite a minor number of events were recorded, 2017 was characterized by large scale attacks (like WannaCry or NotPetya). The Monthly attacks chart, depicted in figure I.4, shows that the

level of activity has dropped slightly from 2016 to 2017, with the exception of September, October and November.

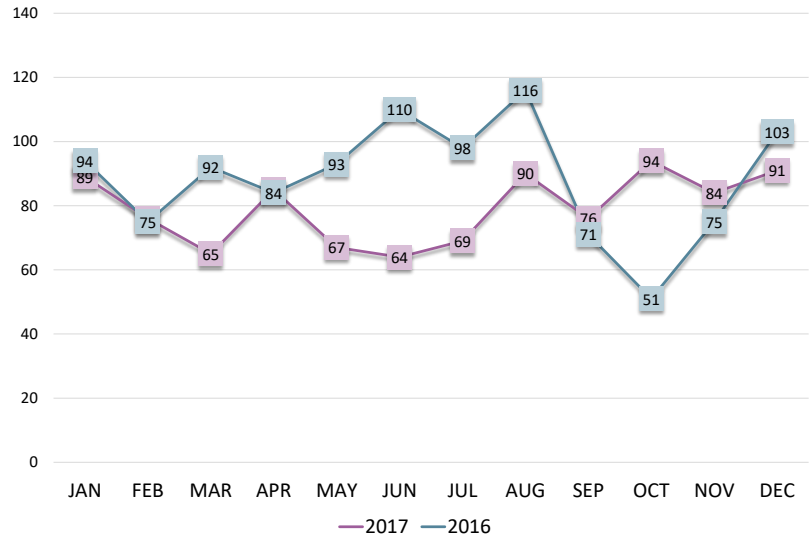


Figure I.4 – Monthly attacks during the years 2016 and 2017 according to HACKMAGEDDON

Malware marked the year 2017. This is the main conclusion of the top 10 attack techniques chart presented in figure I.5. The hijackings (DNS hijacking and account hijacking) were mostly in line with the 2016 results, as like as exploiting vulnerabilities. Targeted attacks reached 15.2% while the fall of hacktivism is one of the possible reasons for the fall of DDOS and SQLi.

I.1.2 Possible solutions

Research and statistics have analyzed many password problems for decades. In fact, passwords are easy to guess, hard to remember, easily stolen and vulnerable to observation and replay attacks [Jobusch & Oldehoeft, 1989, Morris & Thompson, 1979]. Research has invested considerable efforts into alternatives, including biometrics, graphical passwords, hardware tokens, and federated identity. However, text passwords remain the dominant mechanism for authenticating people on computers, and seem likely to remain so for the foreseeable future [Bonneau et al., 2012, Herley & Van Oorschot, 2012]. Considerable efforts have been devoted to the reinforcement of security textual passwords.

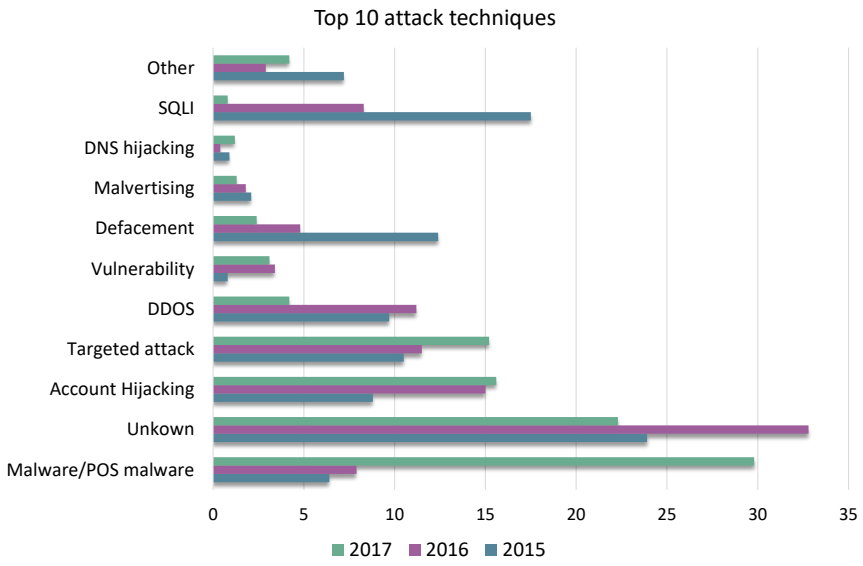


Figure I.5 – Comparison of the most common attack techniques from 2015 to 2017 according to HACKMAGEDDON

I.1.2.1 Password composition rules

Security can be compromised if passwords are easy to guess, although research has consistently shown that users tend to choose simple and easy-to-remember passwords [Shih et al., 2018]. To deal with this problem, password management services often use dialing policies such as the following requirement:

"The password must contain a mixture of uppercase and lowercase letters and at least one digit".

The intended security purpose for password composition rules is to prevent users from choosing easy-to-guess passwords and to guide them to secure password management. A first example of comprehensive guidelines can be found in The Password Management Guide [Brand, 1985]. Nowadays, many rules are automated as a password reinforcement strategy used to regulate the length and composition of passwords. Other password policies include blacklists (Proactive Password Check), password expiration, rate limiting, and lockout policies.

The most common password rules encountered by users include two categories: password reinforcement rules are applied when a password is created and password management rules guide users to secure the management of their passwords.

I.1.2.1.1 Reinforcement rules

Reinforcement rules are usually interested in:

- Length: A password length policy prevents users from choosing passwords that are too short. Password policies typically require a length of at least 6 to 8 characters. There is considerable variability, where some websites may require a shorter length (for example, a 4-digit PIN) while others require longer passwords (for example, at least 8 characters or an exact length).

- Composition: A password composition policy prevents users from choosing passwords that are too simple. It applies rules on the types of characters that can be used. A password composition policy typically requires passwords that contain characters from one or more of the following sets:

    - Uppercase characters
    - Lowercase characters
    - Decimal base digits
    - Non-alphanumeric ASCII characters

  A very simple policy can only insist on a minimal composition (for example, numbers only), while others force a more complex composition (for example, uppercase and lowercase letters, numbers, and special characters). It is suggested that changing complex rules across multiple sites might make it difficult to share passwords between sites [Florêncio & Herley, 2010], but there is little evidence to support that this is the main purpose of a composition policy.

- Blacklist: Some sites prohibit the use of dictionary words because of the susceptibility of selected passwords to password guessing attacks (attackers use lists of dictionaries or probable passwords to guess them) [Habib et al., 2017]. Others apply this rule by prohibiting the use of the most common passwords. For example, users may be prohibited from choosing passwords from a black list of the 1000 most frequently used passwords. Typically, the lists of most commonly used passwords are obtained from disclosed datasets (eg, RockYou).

While the password reinforcement strategies can contribute to their protection against brute-force attacks, it cannot protect users from password capture by malware, social engineering, or physical observation.

I.1.2.1.2 Management rules

Management rules generally advise us to:

- Change it often: An expiration or password aging policy requires users to change their passwords at a fixed time interval (for example, every 90 days). Regular password change has been recommended since a long time by the National Institute of Standards and Technology NIST of U.S. Department of Commerce [Brand, 1985]. The historical security goal is to protect users from the risk of undetected compromise of password over a period of time [Cheswick, 2013, Florêncio et al., 2014]. The more a password is used for authentication purposes, the higher its probability of exposure to hackers. Currently, password expiration is rarely applied in general purpose websites [Florêncio & Herley, 2010], but is commonly used by government and academic institutions.

- Do not reuse: The number of passwords managed by a single user increases with each new account creation. When passwords are reused across multiple accounts, an attacker who compromises a site can use the same password to hack the user's account on another site. Unfortunately, the number of passwords a user needs to remember keeps increasing, which is not easy, knowing that a typical Internet user estimated at 25 separate accounts [Das et al., 2014]. It is suggested that composition policies make password reuse more difficult [Florêncio & Herley, 2010], but they do not directly prevent their reuse. Password policies on different sites are diverse. A unique password policy is only relevant on the site that applies it.

- Do not write it: users are advised not to write their passwords. Difficult passwords generated by password composition rules, pushes users to write them down so they will not be forgotten. The original security intent of this rule is to prevent local attacks from friends, co-workers, family members, or other observers on the spot. Although, plain text passwords should never be stored on unprotected computers with network access, writing passwords copies on papers may not pose a serious security risk [Cheswick, 2013].

- Do not share it with anyone: users are advised not to share their passwords with anyone. The reasons for security seem obvious, but in practice, passwords are often shared with relatives and colleagues. Some security experts [Cheswick, 2013, Stobert & Biddle, 2014] argue that sharing passwords may be appropriate in certain circumstances, such as when recovering accounts or in an emergency.

Studies have shown that password composition policies as well as password indicators (or verbal notifications) help users to choose stronger passwords [Shay et al., 2010, Komanduri et al., 2011, Ur et al., 2012] . However, the harm caused while using an extremely restrictive password policy may be superior to the harm prevented by that policy. Moreover, they may increase the user's annoyance and tiredness [Shay et al., 2010].

A new surprising recommendation of Password Guidelines from NIST consists in removing periodic password change requirements [Grassi et al., 2017]. In addition, the mixture of upper case letters, symbols and numbers is no more needed. NIST indicates that it has been frequently shown that these types of rules often generate worse passwords.

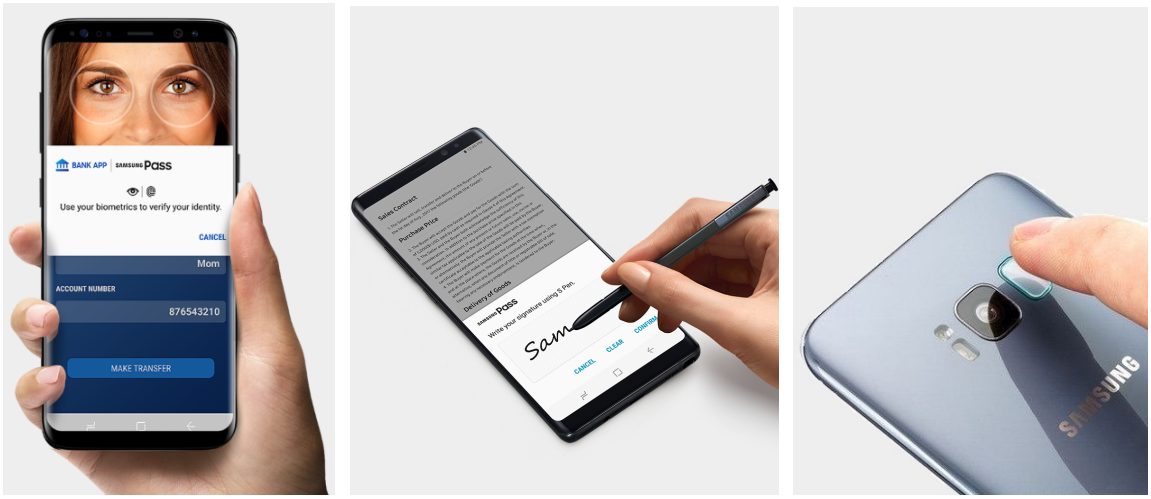
I.1.2.2 Biometric solutions

Other solutions offer the possibility to require biometric data in addition to passwords or to completely replace passwords with biometric digital data. These solutions are nowadays well widespread and even marketed in varieties of novel technological products.

Galaxy S8 or S8+ smartphones, for example, facilitate and secure some of the most sensitive activities such as credit or debit card payment and bank accounts access, with a selection of authentication options, including biometric verification. They scan irises or fingerprints to make purchases with Samsung Pay in-store, then check bank accounts via the Samsung Pass and immediately connect to favorite sites through the web login function, as shown in the figure I.6. These smartphones offer the possibility of unlocking them with fingerprint recognition solution. While using the smartphone, the finger is naturally positioned at the fingerprint sensor placed at the back and it is unlocked in one motion. Increasingly, it seems that in the technological market, passwords will be replaced by the user itself. His face, fingerprints, iris, even heartbeat will authenticate the entry into the digital world.

According to the Biometric Research Group, 650 million people would use their biometric characteristics to unlock their mobile phones by the end of 2015. By 2020, the number of biometric smart-phone users will increase and will reach 2 billion.

As mobile devices such as smartphones and tablets have become ubiquitous, and as personal identification and authentication are increasingly important in the connected world, biometric technologies are becoming an integral part of mobile devices. Biometrics, whether for mobile devices or large stationary systems, usually performs one of two functions: authentication, proof that someone is what he/she claims to be, or identification, determination of who is that person. Almost all identification uses are facing businesses, especially government use cases. Somewhere in the middle, financial institutions offer their users the



(a) Iris (b) Signature (c) Fingerprint

Figure I.6 – Biometric authentication for smartphones

ability to authenticate with online banking systems with their voices or with their iris, instead of entering a Personal Identification Number (PIN).

According to a new report by Tractica, the global market for mobile biometrics will reach $ 3.5 billion by 2024, rising from a base of $ 249 million in 2015, as shown in Figure I.7. The business information firm predicts that cumulative revenues for the 10-year period will total $ 17.5 billion revenue generated by mobile biometric devices as well as revenue from software applications.

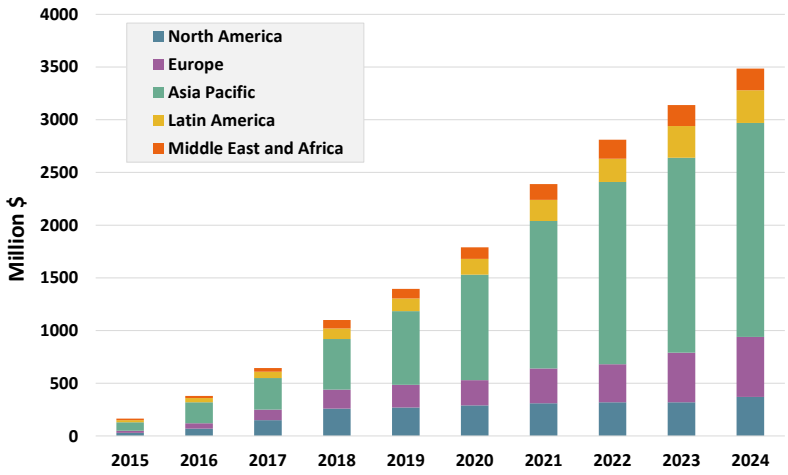


Figure I.7 – Total revenue of mobile biometrics distributed by region around the world

Many basic biometric technologies have been discovered since a long time, but suppliers had no idea to use them, and conditions are now in place to allow mobile devices to exploit

these technologies. This sector can be seen as a classic example of use cases that finally allow to catch up with biometric technology. Most of the revenue in the global market will remain fingerprint authentication on-board, now integrated into smartphones' TouchID. This biometric solution is expected to increase the share of total revenue, as touch-sensitive smartphones becomes more competitive than older smart-phones.

Despite the emergence of biometrics in commercialized applications, keystroke dynamics has been used in few cases by manufacturers and it hasn't been widespread in the technological market until now.

I.2 Thesis objectives

Keystroke dynamics, which will be detailed in the chapter II, is a behavioral biometric solution that authenticates individuals according to their typing manner on the keyboard. Although this modality has proved its efficiency in several scientific researches [Giot et al., 2011b, Pisani et al., 2016, Tsimperidis et al., 2018], it is not yet fully adopted in industrialized applications, unlike other morphological modalities such as fingerprint, iris and face. This is mainly due to the need for several captures during the learning phase to create the model describing the typing rhythm of users. Furthermore, keystroke dynamics suffers from intra-class variation in the user's behavior while manipulating the keyboard.

As part of this thesis, we intend to propose a new efficient approach of logical access control enhanced by keystroke dynamics while addressing the following problems:

- The main idea is to reduce the user's enrollment phase to the strict minimum in order to facilitate the industrialization of keystroke dynamics. It has the advantage of considering a single sample in the enrollment phase to create the reference of the user. Therefore, the proposed method corresponds to the conditions of industrial and operational applications, for which the user enters the password only once when creating an account.

- Proposal of a new update system to overcome the limitations of intra-class variability: This system must work online. It also helps to enrich the description of user's keystroke dynamics by increasing the reference size thanks to the new accepted queries.

- Highlighting a user specific adaptation strategy regarding the difference between users in relation with their typing rhythm behavior.

The aim of this PhD thesis work is to propose contributions on these aspects and to demonstrate the interest of these solutions on the biometric systems developed within the

LATIS and the GREYC research labs. The industrial applications of this work are immediate, especially for authentication on mobile or in the field of payment.

I.3 Main Contributions

The main contributions of this PhD thesis are:

- Proposal of a new adaptation criterion that has the advantage of being individual and adaptive. The *adaptive thresholds* are related to the decision whether to accept the query or not in addition to the decision of updating the reference. Thus, the new adaptation approach consists in updating the reference and the decision threshold at the same time. In fact, we consider that not only the reference must follow the variations of the keystroke dynamics of the user over time, but the decision threshold must be also well chosen. Indeed, a strict threshold does not help to include intra-class variability in the reference. In addition, a very high threshold raises the possibility of including imposter information in the reference. Thus, we used an individual threshold varying from one update session to another so that it is the most appropriate during the use of the system.

- Development of a solution that allows to model the user's keystroke dynamics while minimizing the number of samples serving to the definition of the reference. For this purpose, an enrollment process based on a single sample (the password is typed once during the account creation step) is proposed. The size of each reference of the user increases during the use of the system, to reach a maximum size equal to 10 thanks to the mechanism "double serial". The growing window first serves to enlarge user's gallery in order to capture more intra-class variability. When the maximum size of the reference is reached, the sliding window then takes place and serves to follow the temporal variation of the keystroke dynamics of the user over time. The proposed contribution is an interesting solution because it meets the industrial needs (usability and efficiency).

- Implementation of a GA-KNN verification method to obtain better performance during all adaptation sessions. In fact, the weights obtained thanks to the optimization of the genetic algorithm AG, and associated with the different distances of the KNN classifier, have been useful for minimizing recognition errors. In comparison with previous works, the proposed method showed a great performance improvement.

- Definition of a new update strategy specific to the user category. Generally, a single adaptation mechanism is applied to all users of the authentication process. Although, it has been shown that the performance of biometric systems depends on the user's specificity. Therefore, an update strategy for each category of users has been developed. Then, users are classified based on the Doddington Zoo method. This is a widely used theory for user classification, but it has not been mixed with adaptive strategies for the keystroke dynamics modality. First, the recognition of the user's class according to the animal categories of the Doddington Zoo helps to distinguish the specificity of the user. Then, an adaptive strategy that overcomes the problems of the user class is adopted.

I.4 Thesis outline

In addition to the general introduction, this manuscript is organized around five essential parts, presented as follows:

- Chapter II briefly introduces some notions about biometrics. Then, we recall some generalities on keystroke dynamics as well as the limits of this biometric modality.
- Chapter III describes and compares the different alternatives where the biometric reference is updated through various adaptation strategies. This solution leads to an important and growing area of research, known as *adaptive biometric systems* or *update of biometric models*. This chapter provides an in-depth discussion of several aspects of adaptive biometric systems and evaluation methodology. Some challenges and prospects for future works are also discussed.
- Chapter IV details the first contributions proposed during this thesis. A new update strategy based on a single sample during the enrollment phase is proposed. A new decision criterion for updating "adaptive thresholds" is also considered. In addition to KNN-AG classification, the new "double serial" update mechanism is also introduced.
- Chapter V shows the interest of the new update strategy specific to each category of users. Based on the theory of Doddington Zoo, users are grouped according to their characteristics. Afterwards, suitable adaptation parameters to each group of users are applied.
- Finally, a general conclusion that summarizes this research and provides some potential perspectives.

CHAPTER II

Keystroke dynamics

II.1 Introduction

Password-based authentication is commonly used in our daily lives such as social networks, e-mail, e-commerce and e-banking. Given the increasing number of hacker attacks, the mere use of passwords is not enough to protect personal data. Keystroke dynamics is a promising solution that improves the security of password access by analyzing the user's typing manner. It is a behavioral biometric modality that is increasingly integrated into the logical access security research.

II.2 Biometrics

The term "biometrics" generally refers to the biological, morphological and behavioral characteristics of human beings. However, it is increasingly associated with automated techniques to identify or verify the identity of individuals based on these characteristics. Usually, the identity request of an individual is checked according to what he/she owns (*eg*, keys, a card) or what he/she knows (*eg*, a password, a PIN code). Nevertheless, for biometric recognition, this verification is based on what the individual is, namely the biometric characteristics of the person, such as a fingerprint or a signature.

Biometrics is often considered as one of the most important solutions to security problems involving logical access control (*eg*, a computer, a network, e-commerce, telecommunications); physical access control (*eg*, buildings, airports, etc). According to [Scott Goldfine, 2015], the global biometrics market will grow from $ 2 billion in 2015 to $ 14.9 billion by 2024.

Thereby, a biometric recognition system is generally divided into three major phases as depicted in figure II.1:

- Pre-processing phase: In order to obtain a biometric reference which is the best representative of the user's characteristics, various methods use a data cleaning step that can be manual or automated. This pre-treatment will therefore mainly consist in:
  - do not keep any capture considered as erroneous, in the case of a static authentication (and thus propose a new acquisition to the user);
  - do not keep any capture considered as erroneous, in the case of continuous authentication (which is completely transparent for the user);
  - standardize the data [Hocquet, 2007] by removing the mean and scaling to unit variance or by dividing the features by their standard deviation. It is generally required, especially for machine learning classifiers ;

  - discretize the data [Revett et al., 2006];
  - reduce the size of the space of the captured data.

- Enrollment Phase: Enrollment is the step of registering the person on the authentication system that calculates the user's biometric reference from one or more captures. In the literature, this number of captures is often greater than or equal to 20 for keystroke dynamics [Giot et al., 2011b]. The number of captures required for the creation of the reference can therefore become a hard task for users (especially if they are subject to many mistakes in it). Depending on the studies, either the user's password is requested or one or more texts identical to all users are requested [Gaines et al., 1980]. Generally, the first case is used to set up static authentication mechanisms (at the start of the user's session), while in the second, it is more useful for continuous authentication (the machine constantly checks whether the current user is the connected one).

- Authentication phase: Authentication consists in asking the user to provide his/her identifier and to perform a new capture. Once the capture's data is extracted, the mechanism serves to check if this signature matches the model registered in the system. If both match, the user is authenticated, otherwise he/she is rejected. Most often, this decision criterion is based on a threshold that is set in the system. Figure II.1 shows the different steps of authentication.

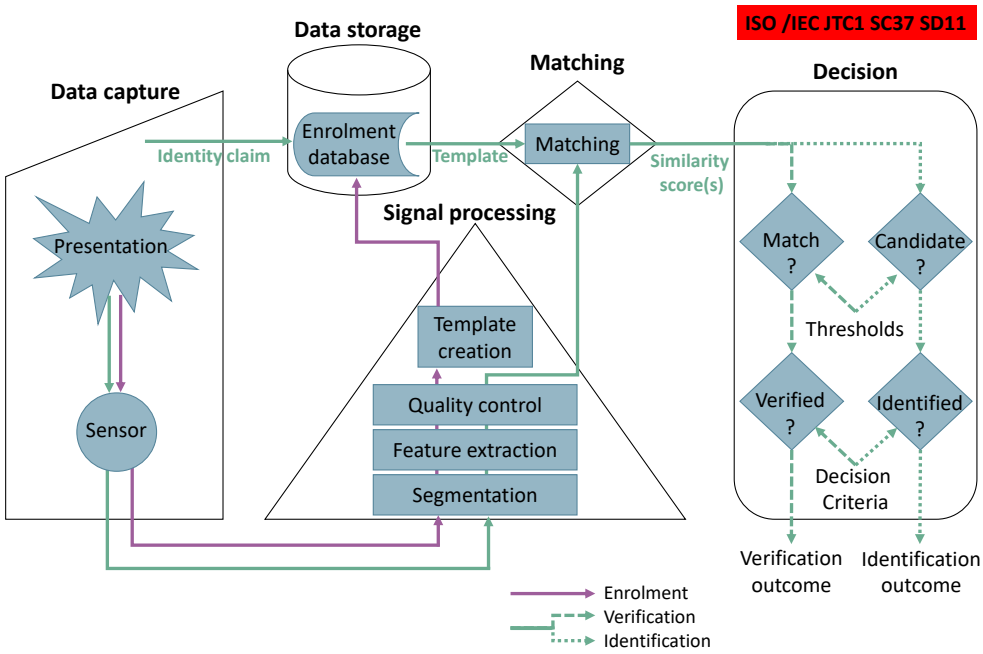


Figure II.1 – Life cycle of a biometric system according to the ISO standard [Bhargav-Spantzel et al., 2007].

II.2.1 Properties of biometric characteristics

To be considered as useful biometric features, any biometric modality must satisfy a number of properties [Jain et al., 2004a], namely:

- Universality: each person must possess the biometric characteristic;
- Distinctiveness: two persons should have sufficiently different characteristics;
- Permanence: the characteristic must be sufficiently invariant (relative to the matching criterion) over a period of time;
- Collectable / Perceptibility: the characteristic can be measured quantitatively.

In addition, biometric authentication systems should also take into consideration [Jain et al., 2011] many aspects:

- Performance: which refers to the achievable recognition accuracy and speed, including the resources needed to achieve the accuracy of the desired recognition at the desired speed as well as the operational and environmental factors that affect accuracy and speed;
- Acceptability: which indicates how much people are willing to accept the use of a particular biometric characteristic as identifier in their daily lives;
- Circumvention: which reflects how it is easy to cheat the system using fraudulent methods.

We note that it is difficult for a biometric authentication system to satisfy all these properties, because they can sometimes be contradictory. Although some properties are specific to the modality (for example, universality, distinctiveness, permanence, perceptibility), others depend on the specific implementation and context (eg performance, circumvention) or culture (for example, acceptability). Some of them are intrinsic to a biometric modality and can not be improved (universality or distinctiveness).

In the next section, a brief description of the most known biometric modalities is presented.

II.2.2 Biometric modalities

The biometric characteristics can be divided into three main classes, namely: morphological, behavioral and biological (see Figure II.2). Thus, a biometric system is essentially a pattern recognition system, which enables personal recognition by determining the authenticity of a specific feature owned by the user:

- The morphological characteristics are related to the body shape: retina, fingerprints (finger, thumb or palm), iris, voice, hand, face, ear, waist, weight, skin and veins;
- The behavioral characteristics are related to a person's behavior: signature dynamics, Handwriting, keystroke dynamics and voices;
- The biological characteristics are linked to the inner part of a living organism: heartbeat, smell, DNA and blood.

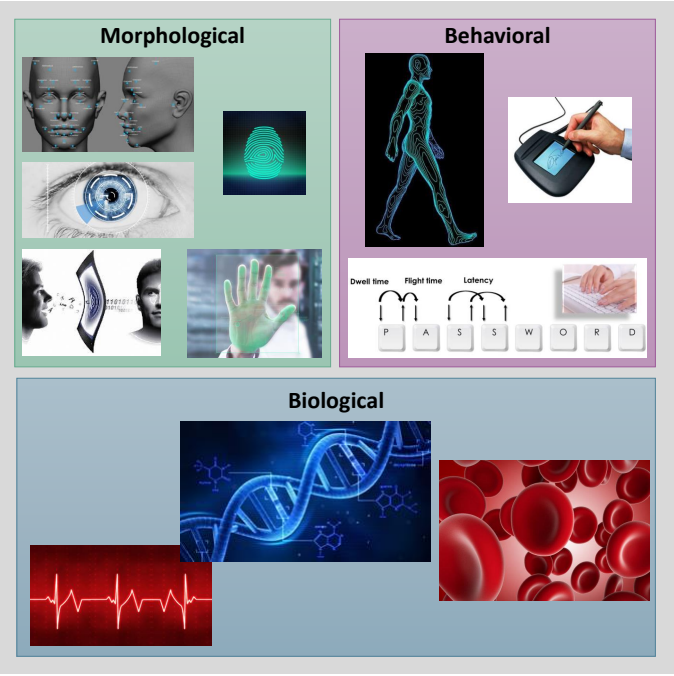


Figure II.2 – Examples of biometric modalities used for user authentication.

II.2.2.1 Morphological modalities

- Fingerprint: Fingerprint recognition is one of the oldest techniques of user recognition. It was developed towards the end of the 19th century by Alphonse Bertillon. The

first automatic authentication system was commercialized in the early 1960s. The fingerprint recognition is based on minutiae matching [Rzouga Haddada, 2017, Vibert, 2017].

- Face: Face recognition is based on the characteristics considered significant as the difference between the eyes, the shape of the mouth, the face round, the position of the ears [El Kissi Ghalleb, 2017].

- Iris: Since 1950, it has been proved that the iris can be used as an authentication factor. It has been proven that the probability of finding two identical irises is less than the inverse of the number of humans who lived on Earth. The treatment of this modality requires that the person be very close to the sensor [Othman, 2016].

- Voice: Voice recognition is not intrusive for the user and does not require physical contact with the sensor. The software recognition can be centralized and the voice transmitted by a network.

  Voice recognition systems are based on unique speech characteristics for each individual. These characteristics are constituted by a combination of behavioral factors (speed, rhythm, etc.) and physiological factors (tone, age, sex, frequency, accent, harmonics, etc.) [Seddik et al., 2004a, Seddik et al., 2004b].

II.2.2.2 Behavioral modalities

- Signature dynamics: This system works with a reader sensor and pencil or pen. This sensor is connected to a computer to control a physical or a logical access.

  Any writing movement of the pen is taken into account but also the movements in high to about 2 cm above the reader. The considered characteristics describing the signature of the user are generally the speed of the signature, variation of the rhythm of the pen, acceleration, pressure, calculation of the distance during which the pen is suspended between two letters, etc. Handwriting, as well, is a personal skill that has been used for the recognition of handwritten note of in postal addresses on envelopes, bank checks, etc [Kacem et al., 2012, Saidani et al., 2015, Kacem & Saïdani, 2017].

- Gait: The research deepens the use of biometrics by creating a recognition system based on the silhouette of the users and their way of walking. This biometric technique offers significant advantages such as remote recognition of the user, without the need for cooperation on his part. Detecting suspicious behavior (via video surveillance), access control to buildings or restricted areas and demographic analysis of a population

in terms of gender and age are some of the possible applications of this technology [Seddik, 2017].

- Keystroke dynamics: The keystroke dynamics is a characteristic of the individual, it is somehow the transposition of a behavior to a vector containing the characteristics of the typing manner of a user [Giot, 2012, Mondal, 2016, Pisani, 2017]. This modality will be developed in the next section in details.

II.2.2.3 Biological modalities

- Electrocardiogram (ECG): The ECG signal is the most common cardiac signal, and refers to the electromagnetic polarization and depolarization of heart muscles over time [Zhang et al., 2017]. It is recorded non-invasively with electrodes attached at the surface of the body. It has been demonstrated that the ECG has sufficient detail for identification.

- Photoplethysmography (PPG): The PPG represents the illumination-based sensing of volumetric changes of blood in the microvascular bed of tissues with every sinus cycle. Measurements can be collected from the fingertip, toe or ear based on a pulse oximeters. Unlike ECG, which measures the electrical activity of the heart, PPG more closely represents the mechanical functioning of the cardiovascular system [Jindal et al., 2016].

- DNA: DNA biometric recognition is highly reliable and secured approach, but hardly applied in real time applications. It is suitable to applications where a high level of security is required. Considering the high cost of DNA analysis and the complexity of sample collection procedure, DNA recognition method is less used in real time applications [Radha et al., 2016].

II.2.2.4 Research of LATIS and GREYC laboratories

This thesis is part of the general framework of major research projects in biometrics conducted both within the team Signal, Image and Document "SID" of the laboratory LATIS and e-payment & biometrics group in the GREYC laboratory.

The SID research axis currently includes 6 projects that focus on the fields of signal, image and document. This PhD thesis is part of the Multimodal Biometrics & Security project, which focuses on data security by designing authentication / verification and watermarking methods for individuals and avatars also.

The e-payment & biometrics group conducts research activities in computer security along two axes with a continuity of theoretical aspects towards the applications, both enriching each other:

- Biometrics: definition and evaluation of biometric systems, protection of biometric data. Several biometric modalities are studied such as keystroke dynamics, fingerprint, face, etc.
- Trust: Applied cryptography, embedded systems, randomness and information protection.

In addition, these two research groups develop numerous software (platforms and mobile applications) to secure user's authentication and different PhD thesis, master projects and graduation projects are conducted in the field.

II.3 Keystroke dynamics

II.3.1 Presentation

The analysis of a user's typing dynamics is a behavioral biometric technique usually described as the way the user manipulates the keyboard. In [Araújo et al., 2005], the authors introduced typing dynamics analysis as a low-cost, non-intrusive authentication method that has a distinct advantage over password authentication: it cannot be lost, forgotten, or stolen. Although the same can be said about morpholgical modalities, they often require expensive and even intrusive equipment to collect biometric data.

Thereby, keystroke dynamics has become a very promising area of research which has been published in several scientific articles, dealing with many topics such as the choice of extracted characteristics, classification methods, the combination with other biometric modalities (multi-modality) as shown in Figure II.3.

This behavioral biometric modality combines the verification of the syntactic accuracy of the password with the conformity with the behavior of the legitimate user, his/her typing rhythm on the keyboard. Various studies have been conducted to highlight this modality. The two main coexisting families are:

- *Static text authentication* where the user always types the same text. This text is usually a pre-defined password. It may be common for all users (a passphrase), or it may be a user specific password. This is the most utilized category in the literature [Hocquet et al., 2007, Idrus et al., 2014, Killourhy & Maxion, 2010, Giot et al., 2011b, Tsimperidis et al., 2018].

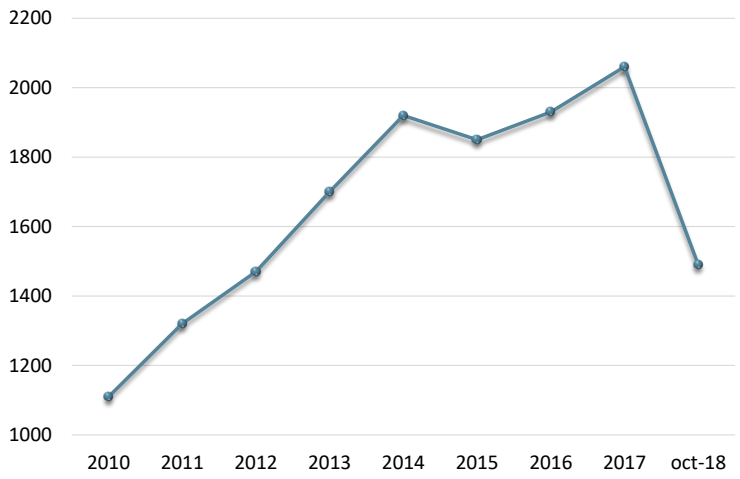


Figure II.3 – Number of scientific publications per year focusing on keystroke dynamics

- *Free text authentication* where the user does not always enter the same text [Bours & Barghouthi, 2009, Xi et al., 2011, Messerman et al., 2011, Pinto et al., 2014, Bours & Mondal, 2015, Nilsen, 2018, Aljohani et al., 2018]. There may be continuous authentication, which constantly checks the identity of the user. Challenge-based authentication should be considered in some applications. It asks the user to a enter text he/she does not know in advance as a challenge to avoid the replay attack. The server needs to verify also whether the user typed the assigned text.

In both cases, the extracted characteristics describing the user's typing manner are practically the same. They are detailed in the next section.

II.3.2 Extracted features

The characteristics to describe a user's keystroke dynamics are generally timing events acquired by the Operating system (OS). Other studies consider video and sound data that are collected by recording users during data collection [Vural et al., 2014].

Regardless the typed text, the keyboard provides the times when each key is pressed and released. From these basic data, the characteristics are extracted and used as input for the classification algorithm. To describe the keystroke dynamics of one user, researches are frequently interested in temporal information extracted from digraph transition times. In this manuscript, we have adopted the following notation to represent extracted features. The figure II.4 shows these characteristics in a graphical way, wherein the arrows downwardly and upwardly respectively denote the moments of pressure and release of each key password:

- PP: time difference between the press events of two successive keys;
- RR: latency between the release events of two successive keys;
- RP: time duration between a one-key release event and its following key press event. It is also called "flight time";
- PR: time duration between a one-key press event and its following key release event;
- Dwell: time duration between a one-key press event and its release event. This characteristic represents the time that the hold key is pressed and it is also called by some authors "dwell time".

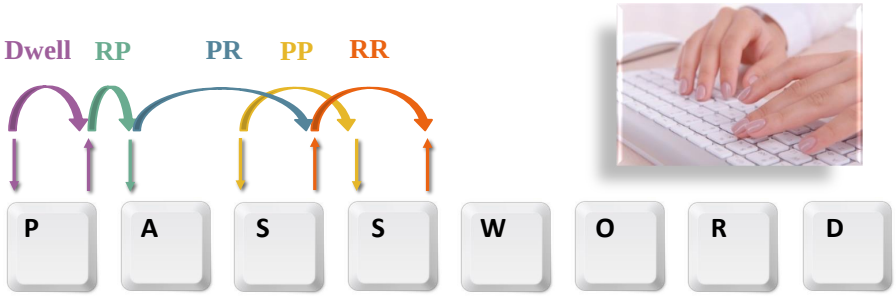


Figure II.4 – Characteristics of the keystroke dynamics.

The feature vector is then generated based on these characteristics. An example of a feature vector for an expression of four keys is shown in Figure II.5. A summary of the features used in some research works in the literature is presented in the table II.1. From the data in this table, we generated the histogram represented on the Figure II.6. It is clear that, the characteristics Dwell (dwell time) and RP (flight time) are the most used ones.

Another feature that can be used is the pressure on the keys [Chang et al., 2012, Elftmann, 2006], but extracting this feature requires adding a specialized hardware. However, with the increasing availability of touchscreen devices, the costs of using this feature may decrease over time. In [Chang et al., 2012], the pressure on a touch screen smartphone was evaluated in a keystroke dynamics scenario. Error rates decreased from 12.2% to 6.9% when pressure was also taken into consideration.

Studies have applied pre-processing steps to improve the quality of the acquired characteristics and subsequently improve the recognition performance. In [Montalvao et al., 2006], an equalization process on the feature vector is applied. Authors argue that this transformation can highlight important aspects of the feature vector, as has been observed in other areas,

Table II.1 – Extracted features of keystroke dynamics modality

Reference	Extracted features
[Muliono et al., 2018]	Dwell, RP, PP
[Bours & Ellingsen, 2018]	PP, RR, RP, PR
[Pisani et al., 2017]	RR, PP, RP, PR
[Monaro et al., 2017]	Dwell, RR, PP, RP, PR and total typing time
[Idrus et al., 2014]	RR, PP, RP, PR
[Giot et al., 2012a]	RR, PP, RP, Dwell
[Chang et al., 2012]	Dwell, RP, PP, pressure
[Killourhy & Maxion, 2012]	Dwell, PP
[Giot et al., 2011b]	RR, PP, RP, PR
[Killourhy & Maxion, 2010]	Dwell, PP, RP
	Dwell, PP
	Dwell, RP
[Giot et al., 2009b]	PP, RR, RP, PR and total typing time
[Killourhy & Maxion, 2008]	Dwell, RP
[Hosseinzadeh & Krishnan, 2008]	Dwell
	PP
	RR
	RR, PP
	Dwell, PP
	Dwell, RR
	Dwell, RR, PP
[Rodrigues et al., 2006]	RP, Dwell
	RP, Dwell, RR, PP
[Bartlow & Cukic, 2006]	Dwell, RP (average, standard deviation, sum, minimum and maximum), including the Shift key
[Montalvao et al., 2006]	PP
	PP with equalization
[Chang, 2006]	Dwell, RP
[Montalvão Filho & Freire, 2006]	PP
	PP with equalization
[Gunetti & Picardi, 2005]	Dwell, RP
[Yu & Cho, 2003]	Dwell, RP
[Monrose & Rubin, 2000]	Dwell, RP

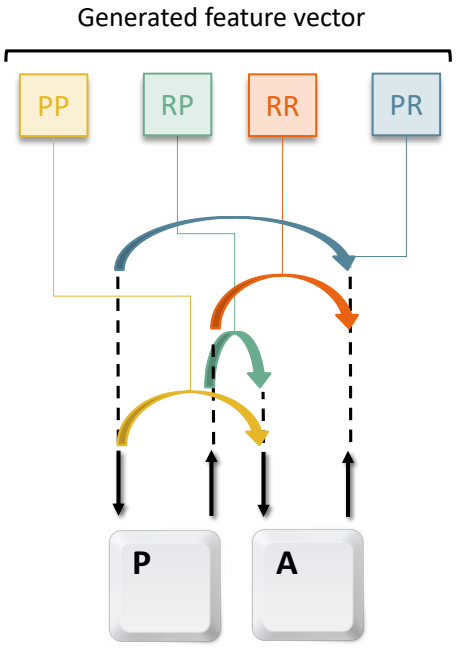


Figure II.5 – Example of a feature vector

such as digital communications and image processing. According to the reported results, the application of this equalization has improved performances (lower error rate) obtained by several previous research works.

Moreover, in [Giot et al., 2011b, Giot et al., 2009b] authors evaluated the use of discretization of feature vectors. Each value of the feature vector is discretized in five ranges. The discretized data are then classified by a two-class SVM, using both negative and positive samples for training. According to the authors, the application of SVM with this discretization has obtained lower error rates than other approaches seen in the literature (for example, neural networks and distance-based classifiers).

In [Hosseinzadeh & Krishnan, 2008], authors performed a comparative analysis of seven sets of characteristics. All combinations using Dwell, PP and RR were considered. The best performance was achieved by the Dwell and RR characteristics. However, the RP functionality has not been taken into account in their analysis. RP is one of the most used features in previous works, as shown by the Figure II.6.

Another study on extracted characteristics was carried out by [Bartlow & Cukic, 2006]. In addition to considering the key "characters", this study was also interested in the shift key. In passwords that contain a mix of lowercase and uppercase letters, the shift key is obviously

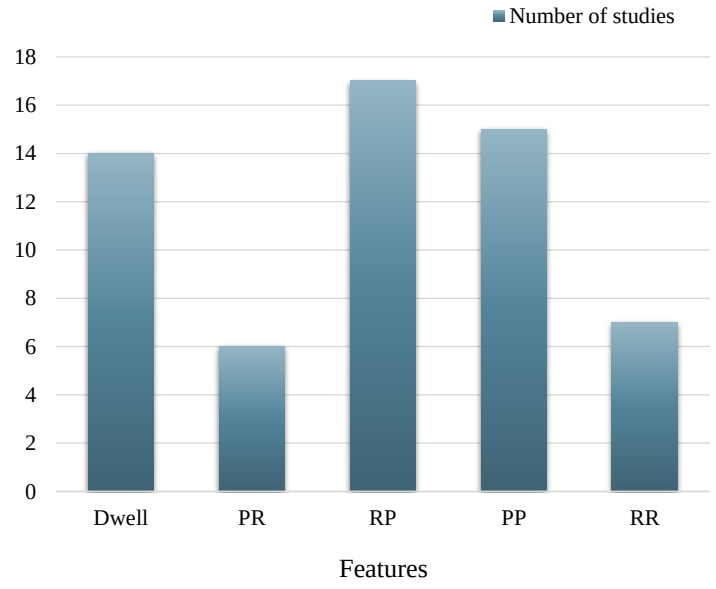


Figure II.6 – Number of research works that used each characteristic

used. Therefore, the analysis of the shift key can be an additional factor to rank the users. According to their tests, the analysis of the shift key reduces the error rates of the classifier.

An important factor concerning the keystroke dynamics modality is the resolution of the captured data. In the MS Windows operating system, for example, the notification of keyboard events, such as pressing and releasing keys, does not distinguish differences less than 15.625 ms. In [Killourhy & Maxion, 2008], the effect of different resolutions was evaluated. This evaluation used an external device with a resolution of $100\mu s$. High resolution data was then used to derive samples at lower resolution. As expected, the higher resolution data imply better classification accuracy. Low resolutions (eg, 100 ms) resulted in error rates of 50%, which is a very poor recognition performance.

Recently in [Monaro et al., 2017], the authors asked the users to complete the required fields with their real autobiographical information (identity, birth, residence, education, interests, etc). Then, using keystroke dynamics analysis they detect users providing false personal information during the authentication process to an online service. As previous work on user authentication considered that a timing resolution from $0.1s$ to $1\mu s$ is sufficient to capture typing characteristics, data were time stamped and measured up to microseconds (μs) precision.

II.3.3 Biometric databases

Most studies in biometric field dealing with keystroke dynamics modality require a biometric database for the validation of the obtained results. Some datasets identified in the literature are shown in Table II.2. The availability of suitable public datasets for the evaluation of biometric systems is generally limited. A possible reason is the intrinsic difficulty to acquire data for such kind of study as these datasets need to contain several samples per user. Ideally, they should be obtained at different acquisition sessions, either with different acquisition conditions or separated by a certain amount of time, to justify the use of adaptive biometric systems. But for keystroke dynamics, public databases are a little more available. For our experiments, we chose three datasets, among the most widely used in the literature, to validate the proposed contributions:

- GREYC 2009 [Giot et al., 2009a]: This database was developed within the GREYC Laboratory. One hundred and thirty-three users participated in the creation of this database and typed the same password "greyc laboratory". Only 100 of them participated in five acquisition sessions during two months and provided 60 samples per user. These samples were focused on in our experiments. This database were chosen to compare our results with those of the experiments in [Giot et al., 2011b]. The database contains both raw and extracted data. The extracted features are : PP, RR, RP and PR.

- GREYC-Web [Giot et al., 2012a]: For this database, 118 users were involved in its creation and typed the same password "SÉSAME". Only 45 among them participated in five sessions and provided 60 patterns. These users were the subject of our experiment. Another advantage of this dataset is that it contains two types of biometric data: (i) imposed password common to all users, as for 100% of other public databases, (ii) passwords chosen by the user with impostor attacks provided by other users, unlike other public databases. This database contains both raw and extracted data which are : PP, RR, RP and PR.

- CMU [Killourhy & Maxion, 2010]: This database includes data of 51 users. They typed the same password 400 times during eight acquisition sessions. The time between each session is at least one day, but the average value is not specified (it can be expected to be different depending on the users). This is the public dataset with the greatest number of samples per user, but many acquisitions are made over a relatively short period (50 acquisitions per session).The defined password was ".tie5Roanl". The database contains only the extracted Dwell, PP and RP characteristics. We opted for this database because it was frequently used in the literature.

Table II.2 – Datasets used in the evaluation of keystroke dynamics biometric modality.

Datasets	# Users	Period/Sessions
GREYC [Giot et al., 2009a]	100	2 months (5 sessions)
GREYC-Web [Giot et al., 2012a]	118	more than 1 year
CMU [Killourhy & Maxion, 2010]	51	8 sessions

By analyzing the existing literature, as shown in Table II.2, we found that the number of users and the time period or sessions among the datasets differ significantly. While it is generally true that a higher number of users and longer sessions can result in a more reliable expected performance, the variability in the nature and context of experiments means that it is extremely difficult to compare different adaptation techniques.

II.3.4 Classification algorithms

A number of algorithms have been used to classify users depending on their keystroke dynamics modality. Table II.3 shows the algorithms studied in some selected publications. It is important to note that, apart from the machine-learning algorithms known in the literature, such as Support Vector Machines (SVM) [Giot et al., 2009b] and Nearest Neighbor [Killourhy & Maxion, 2008], some authors have proposed new algorithms [Monrose & Rubin, 2000, Gunetti & Picardi, 2005]. These new algorithms have also been used in comparisons made by subsequent research [Montalvao et al., 2006].

The statistical classifiers have been deeply used. They are based on calculating statistical characteristics from training samples (e.g. mean, median and standard deviation) and comparing them to those of the new introduced query using various distance metrics. Three main statistical classifiers have been used [Hocquet et al., 2007, Bleha et al., 1990a, Revett et al., 2006, Boechat et al., 2007]in the litterature.

The use of static and free text has been tested in [Monrose & Rubin, 2000]. The authors conducted different experiments to validate the idea of classifying users according to their typing rhythm based on various distances and probability tests. Their experiments validate the approach, reaching a precision rate of 92.14%.

As discussed in previous studies [Killourhy & Maxion, 2010, Giot et al., 2009b], the number of training samples can affect the performance of the classifier. In general, the greater their representativity, the higher the classification accuracy. In [Chang, 2006], a method for generating new learning samples based on the genuine user has been proposed. The samples are generated using time domain resampling and using the Discrete Wavelet Transform

Table II.3 – Classifiers used in the keystroke dynamics modality.

Reference	Classifieur
[Muliono et al., 2018]	Deep learning SVM (Linear, RBF and Polynomial)
[Pisani et al., 2016]	K Nearest Neighbor
[Giot et al., 2012a]	Based on Gaussian distribution [23]
[Chang et al., 2012]	Statistical [Boechat et al., 2007]
[Killourhy & Maxion, 2012]	Statistical Disorder-based
[Giot et al., 2011b]	SVM Statistical Neural network Distance-based classifier
[Killourhy & Maxion, 2010]	Nearest neighbour Outlier count (z-score) Manhattan distance
[Giot et al., 2009b]	SVM Statistical Classifier based on Euclidean distance Classifier based on Hamming distance
[Killourhy & Maxion, 2008]	Nearest neighbour Neural network Mean-based classifier
[Hosseinzadeh & Krishnan, 2008]	Gaussian Mixture Model (GMM) + Leave one out method
[Montalvao et al., 2006]	[Bleha et al., 1990b] [Monrose & Rubin, 2000] [Gunetti & Picardi, 2005]
[Montalvão Filho & Freire, 2006]	[Bleha et al., 1990b] [Monrose & Rubin, 2000] 1D-Histogram and 2D-Histogram
[Rodrigues et al., 2006]	Hidden Markov Model (HMM) Statistical
[Bartlow & Cukic, 2006]	Random Forests
[Chang, 2006]	Tree-based with Euclidean distance
[Gunetti & Picardi, 2005]	Proposed Methods: R Measure and A Measure
[Yu & Cho, 2003]	SVM 2-layer and 4-layer Auto Associative Multi-layer Perceptron (AAMLP)
[Bleha et al., 1990b]	Euclidean distance Weighted and non-weighted probability Bayes

(DWT). Although this method generates more samples, an unresolved question is whether these new samples actually imply better representativeness.

The use of numeric keypads has been analyzed by [Killourhy & Maxion, 2008] since quite a long time. An advantage of using digital keyboards is that it would be easier to implement typing dynamics technology in mobile devices, such as cell phones, which typically have only one keypad nowadays. The authors conducted experiments using eight passwords, obtaining an ERR of 3.6%. As mobile touch devices have become ubiquitous everyday tools, research is increasingly interested in user's interaction with touchscreen keypads [Buschek, 2018, Li & Bours, 2018].

New detectors were tested in [Yu & Cho, 2003], especially an autoassociative multilayer percetron (AAMLP) and a one-class support vector machine (one-class SVM). According to their experiments, error rates were similar for both novelty detectors. However, the one-class SVM was more efficient in terms of computational resources usage.

In addition, studies using Neural Networks (NN) have been frequently applied to keystroke dynamics since 1993 [Brown & Rogers, 1993, Bleha & Obaidat, 1993, Anagun, 2006, Ahmed & Traore, 2014]. NN have the disadvantages of requiring a huge number of labeled samples (from genuine and impostor users) in order to create a reference template. Moreover, in this case, parameters setting is rather complex. The efficiency of Support Vector Machine (SVM) classifiers have been also tested [Sang et al., 2004, Giot et al., 2011b]. They have been used in the context of either one-class or two-class classification (where impostor attacks were considered). For one-class classification, the authors proposed in [Yu & Cho, 2004] the Genetic Algorithm (GA)-SVM wrapper approach. They improved the SVM classification by adding the GA to perform features selection. Accordingly, the created user's model demonstrated a better performance, but the number of samples used to create the reference was large as well (equal to 50).

Many other classifiers have been used in the literature for keystroke dynamics authentication systems, such as the Bayesian classification [Bleha et al., 1990a], the Hidden Markov Model [Rodrigues et al., 2005] or the K Nearest Neighbor (KNN) classifier. For example, the authors in [Pisani et al., 2016] opted for the KNN classifier to distinguish genuine samples from impostor ones, and then to create two galleries: a positive one, to save samples classified as genuine, and a negative one, to collect samples classified as an impostor. The positive gallery was composed of 40 samples captured during the enrollment phase.

II.3.5 Performance metrics

Most studies reported in the literature that evaluate keystroke dynamics biometric systems use the same metrics as those used to assess the majority of biometric modalities [Himaga &

Kou, 2008, Precise Biometrics, 2014, Poh et al., 2014]. In this section, we first discuss the standard metrics for the evaluation of biometric systems.

- *FNMR (False Non-Match Rate)*: the rate of genuine *attempts* that were wrongly classified as impostor. In order to report the global FNMR, the average from all users can be computed. Another approach is to simply compute the metric considering the amount of genuine queries from all users at the same time. Note that when the number of genuine queries is different among the users, these two methods to compute the global FNMR can result in different values. The first method gives the same weight to each user, while the second method gives more weight to those users which contain a higher number of genuine queries.

  A related metric is FRR (False Rejection Rate), which his similar to FNMR, but also considers the FTA (Failure to Acquire Rate), is shown in Equation (II.1). FTA measures the rate in which a biometric system fails to obtain a biometric sample.

$$FRR = FTA + FNMR \times (1 - FTA) \tag{II.1}$$

- *FMR (False Match Rate)*: rate of impostor *attempts* that were wrongly classified as genuine. The global FMR, generally measures the average FMR from all users. Another approach is to simply compute the metric considering the amount of impostor queries from all users at the same time. Note that when the number of impostor queries is different among the users, these two methods to compute the global FMR can result in different values. In some evaluation methodologies, the number of impostor queries is a function of the amount of genuine queries. In the first method, each user has the same weight. In the second method, users which contain a higher number of impostor queries receive a higher weight).

  A related metric is FAR (False Acceptance Rate), which has almost the same definition of FMR, similarly to the case of FNMR/FRR. FAR also considers the FTA, as shown in Equation (II.2).

$$FAR = FMR \times (1 - FTA) \tag{II.2}$$

- *HTER (Half Total Error) and balanced accuracy*: HTER is defined by Equation (II.3) as the average between FNMR and FMR. This metric combines the results from both FNMR and FMR in a single value, making the performance evaluation simpler. This measure can also be defined using the balanced accuracy *BAcc* [Masso & Vaisman, 2010], defined in Equation (II.4).

$$HTER = \frac{FNMR + FMR}{2} \tag{II.3}$$

$$BAcc = 1 - HTER \tag{II.4}$$

- *EER (Equal Error Rate)*: it is the value when FNMR is equal to FMR. This metric can be seen as a particular case of HTER, when $FMR = FNMR$.
- Area Under Curve (AUC): It is the measure of the area of the surface below the Detection Error Trade-of (DET) curve (FMR versus FNMR).

These metrics can also be obtained over time. The study in [Rattani et al., 2011] claims to be the first one in the area to compute results over time, instead of just reporting it globally. Later, a plot to report performance metrics over time in the context of a biometric data stream has become common for several works.

A metric specific to describe the performances of continuous authentication for keystroke and mouse dynamics is proposed in [Bours & Mondal, 2015]. Indeed, two new metrics Average Number of Impostor Actions ANIA and Average Number of Genuine Actions ANGA respectively comparable to FMR and FNMR in a static authentication. In fact, for an impostor user i, when tested against the template of genuine user g and is locked out k times after $N_1, N_2, \ldots, N_k$ actions, then the $ANIA_g^i$ is defined as :

$$ANIA_g^i = \frac{1}{k} \sum_{j=1}^{k} N_k \tag{II.5}$$

An ANIA of the attacks of $M - 1$ different imposters (all users except the genuine one) against each genuine user g is then equal to:

$$ANIA_g = \frac{1}{M-1} \sum_{i=1}^{M-1} ANIA_g^i \tag{II.6}$$

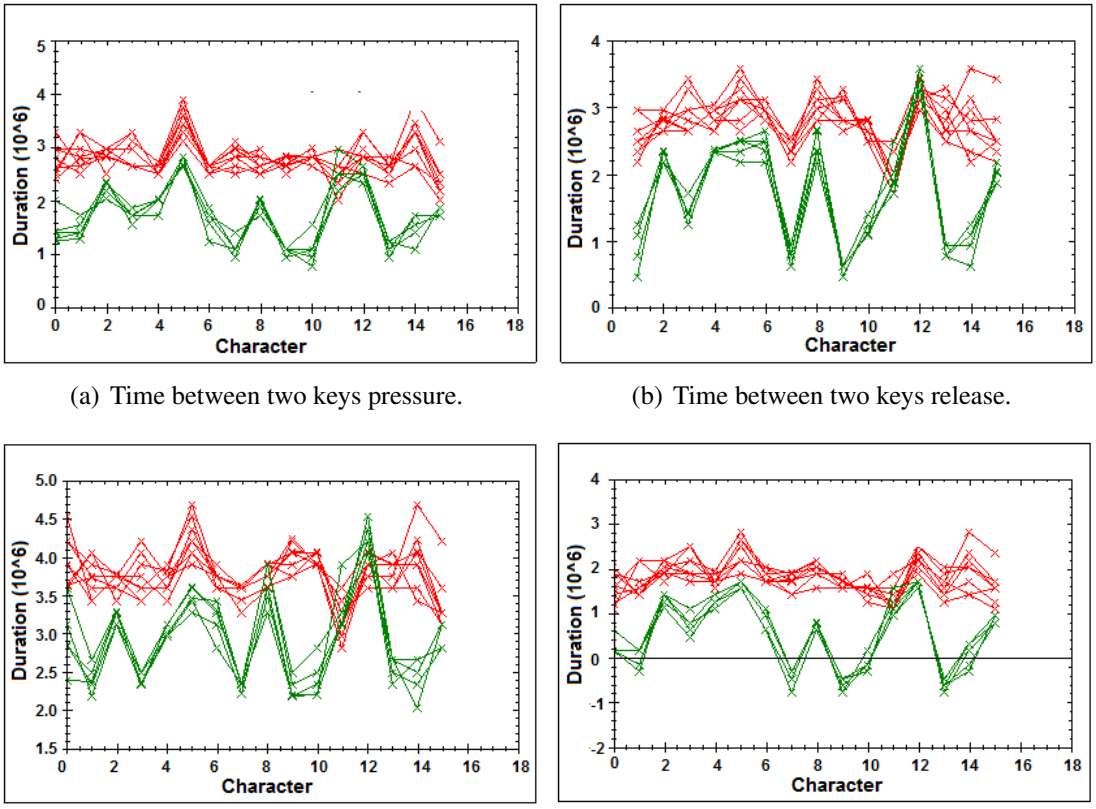
In general, the ANIA against all the users of the system equals to:

$$ANIA = \frac{1}{M} \sum_{g=1}^{M} ANIA_g \tag{II.7}$$

Accordingly the $ANGA_g$ of a user g is calculated when his own query is tested against his own template. The average of the $ANGA_g$ values over all users is then defined as the $ANGA$ of all the system.

II.3.6 Recognition errors

This section describes some sources of variability in the biometric features over time, which can result in *template aging*, motivating the need for performing adaptation in biometric systems. In the machine learning literature, the term *concept drift* is often used with the meaning of change in the profile of the data distribution. In the context of biometrics, the drift is caused by a plethora of factors or sources of variability. Moreover, previous experimental results have shown that this variability can be different depending on the user [Poh et al., 2015a, Pisani et al., 2015b, Rattani et al., 2009]. As illustrated in Figure II.7, one month later (green curves) the keystroke dynamics of the user is quite different from the initial one (red curves).



(a) Time between two keys pressure.

(b) Time between two keys release.

(c) Time between one pressure and one release.

(d) Time between one release and one pressure.

Figure II.7 – Intra-class variability of a user after one month

This variability may be due to different errors like :

- *Errors related to enrollment/changing conditions*: The model may not represent the user's characteristics when limited amount of samples is available during the enrollment stage. Besides, the acquisition conditions during the recognition phase may not be

the same as those of the enrollment phase. For example, aspects like illumination, humidity, noise, movement and portability of the device can vary between enrollment and recognition phases [Poh et al., 2009b]. Another source is the use of devices with distinct characteristics for enrollment and recognition (*cross-device matching*). This can occur, for example, in face recognition, when the quality of images produced by high-resolution cameras and web-cameras can be very different [Poh et al., 2010a]. In fingerprint recognition, this can occur when matching two samples collected using thermal and optical fingerprint sensors.

Self occlusions (e.g., make up) and occlusions due to the use of accessories (*e.g.*, body-piercing ornament and jewelries, or glasses) can also introduce intra-class variability. Furthermore, the interactions between a user and each sensor can be different. In keystroke dynamics, to type on different keyboard layouts can produce different keystroke dynamics [Jain et al., 2016]. Emotion and health can also impact the recognition performance of a biometric system, especially in the behavioral modalities. Emotional states, such as happiness, anger and stress can change the speech sound.

- *Errors related to time/aging*: Both physical and behavioral biometric modalities are subject to changes related with time/aging. Physical modalities are subject to injuries, wrinkles, speckles, weight loss and gain. Moreover, illnesses and/or their associated treatments can also permanently alter samples in the speech and fingerprint modalities. Behavioral modalities features are subject to habit changes. As time elapses, the user may become familiar with the system, changing how he/she use the system. For example, in keystroke dynamics recognition, the users may change their typing rhythm over time [Montalvão et al., 2015].

The sources of intra-class variability can increase false non-match errors. When genuine users are increasingly rejected by the system, they can become annoyed, negatively impacting the usability of the system. Manually re-enrollment of the users periodically can be a solution, although it is costly. Two other alternatives to decrease the impact of intra-class variability are using multi-modal biometric systems and soft biometrics.

Multi-modal biometric systems use multiple biometrics [Ross et al., 2006] to reduce the overall system error can occur at different levels (sensor, characteristics, score, rank or decision). Despite its efficiency, the configuration of system parameters become more complex, increasing the cost of the recognition system. The fusion can also be inconvenient to the user, since it can increase the overall authentication time and require the user to learn to use several sensors.

Different from classical biometrics, soft biometrics [Jain et al., 2004b] can improve biometric system performance by using characteristics that are not unique and not permanent to sufficiently differentiate between two individuals, though they can support the recognition decision. Examples are gender, age, ethnicity, skin, hair color or educational level [Idrus et al., 2014, Tsimperidis et al., 2018].

Both multi-modal and soft biometric systems can decrease the impact of template aging by increasing the recognition accuracy. However, they are also subject to template aging. For example, in multi-modal system combining fingerprint and face recognition, when biometric features for both biometric modalities change, the recognition performance can degrade over time.

This thesis focuses on adaptive biometric systems, which are able to automatically adapt the biometric reference over time. They are sometimes referred to as *template update* in the literature. Next sections of this manuscript focus on adaptation strategies.

II.4 Conclusion

Biometric systems, especially those based on a behavior analysis are able to differentiate the identity of an individual according to what he / she is doing. They are a promising alternative to limit the identity usurpation. Among the characteristics to be analyzed in order to define the user's behavior, this work focuses on keystroke dynamics as a biometric modality.

In this chapter, the main goal was to identify the state of the art of keystroke dynamics modality. To accomplish this task, we identified the advantages and disadvantages of using typing dynamics, characteristics extracted from typing data, classification algorithms, and performance metrics.

The biometric features used for users recognition should satisfy certain properties, as discussed at the beginning of this chapter. However, recent studies have shown that *permanence* is not satisfied for several biometric modalities [Rattani, 2010, Giot et al., 2011b, Pisani, 2017]. This is due to several reasons, including aging and changing conditions, as mentioned in section II.3.6. In order to solve this problem, adaptive biometric systems have been proposed. This is a relatively new field of study in biometrics.

CHAPTER III

Strategies to adapt the Biometric Reference

III.1 Introduction

With the widespread of computing and mobile devices, authentication using biometric traits like face, iris, voice and keystroke dynamics has received greater attention in logical access control like web services authentication. Although biometric systems usually provide good solutions, the recognition performance tends to be affected over time due to changing conditions and aging of the biometric data, resulting in intra-class variability. Adaptive biometric systems, which adapt the biometric reference over time, have been proposed to deal with such intra-class variability.

This chapter provides the most up-to-date and complete discussion on adaptive biometric systems we are aware of, including formalization, terminology, adaptation strategies and open challenges. This work is a part of a survey that was achieved in collaboration with P.H. Pisani, R. Giot, N. Poh and A.C.P.L.F. De Carvalho.

III.2 Biometric systems

III.2.1 Generalities

A *biometric system* is a pattern recognition system that acquires a *biometric query sample* of the *claimant* and extracts its *biometric features* to compare them with a previously stored *biometric reference* corresponding to the *claimed identity* in a biometric database [Jain et al., 2004a]. A *biometric reference* is also known as a *model* or *template*.

Many studies have shown that biometric features may change over time [Roli et al., 2008] and, consequently, the biometric reference may no longer represent the biometric features of the user. This phenomenon is known as *template aging* [Jain et al., 2016]. As a result, the recognition performance of the biometric system can degrade over time. An *adaptive biometric system* adapts the user reference to deal with template aging [Roli et al., 2008, Poh et al., 2012].

This section presents the terminology adopted in this manuscript and describes the main parameters of an adaptive system.

III.2.2 Terminology

A biometric system has two main phases: *enrollment* and *test/recognition*. In the enrollment phase, defined by Equation (III.1), the system receives a set of enrollment samples $\mathcal{E}_j$ for each user $j \in \mathcal{J}$ and outputs its biometric reference ref_j, where $\mathcal{J}$ is the set of user

indexes registered in the biometric system. The enrollment is performed for all registered users and each biometric reference is stored in a biometric database $\mathcal{R} = \{ref_j \mid j \in \mathcal{J}\}$.

$$ref_j \leftarrow enroll\left(\mathcal{E}_j\right) \tag{III.1}$$

Another phase is the *test/recognition*. In this phase, the system receives a biometric query sample $\mathbf{q}$ and returns the identity label of the recognized user. A query is a biometric sample acquired to perform recognition. The *test/recognition* can operate in *verification* or *identification* modes [Jain et al., 2016].

In the *verification* mode, defined in Equation (III.2), a query $\mathbf{q}$ is compared to the biometric reference ref_j of a claimed user index j using a set of parameters θ_j^{verify}. The output is obtained from a *classification algorithm*, which returns the predicted label $label^p$ for the biometric query: *genuine* or *impostor*. The set θ_j^{verify} refers to the parameters adopted for the classification algorithm. Some implementations output a *score* from the comparison of a query $\mathbf{q}$ to the biometric reference ref_j and afterwards return the class label by comparing this *score* to a *decision threshold* value. In this case, the *decision threshold* would be an element in the set of parameters θ_j^{verify}. Other classification algorithms may need other parameters, such as the kernel parameters for a support vector machine [Schölkopf et al., 2001].

$$label^p \leftarrow test.verify\left(ref_j, \mathbf{q}, \theta_j^{verify}\right) \tag{III.2}$$

In the *identification* mode, defined in Equation (III.3), a query $\mathbf{q}$ is presented to the biometric system, which outputs a set of user indexes $\mathcal{U}_{id}$ using the set of parameters $\theta_j^{identify}$, such as $\mathcal{U}_{id} \subseteq \mathcal{J}$. The set $\theta_j^{identify}$ refers to the parameters of the classification algorithm used, as in the case of the verification mode (e.g. *decision threshold*). Note that $\mathcal{U}_{id}$ can be a null set {} when the query is classified as an *impostor*.

$$\mathcal{U}_{id} \leftarrow test.identify\left(\mathcal{R}, \mathbf{q}, \theta_j^{identify}\right) \tag{III.3}$$

In addition to the previously discussed *enrollment* and *test/recognition* phases, which are applicable to any conventional biometric system, an adaptive biometric system can also operate in *adaptation* mode. In the adaptation phase, the *adapt* process, as specified in Equation (III.4), adapts the biometric reference $ref_{j_{(t)}}$ using a set of biometric samples for adaptation $\mathcal{A}$, along with a set of adaptation parameters θ_j^{adapt}. The output of the adaptation process is the *adapted biometric reference* $ref_{j_{(t+1)}}$.

$$ref_{j_{(t+1)}} \leftarrow adapt\left(ref_{j_{(t)}}, \mathcal{A}, \theta_j^{adapt}\right) \tag{III.4}$$

The set of samples used for adaptation, $\mathcal{A}$, is collected during the system operation. Usually, it only contains samples classified as genuine by the test/recognition process. As discussed later, some studies only include samples classified as genuine with high confidence in this set [Roli & Marcialis, 2006, Giot et al., 2011a]. Thus, an additional *adaptation threshold* that is more stringent than the *decision threshold* can be used. In this case, the *adaptation threshold* would be an element in the set of parameters for adaptation θ_j^{adapt}. Depending on the adaptation strategy, there may be different parameters, as it will be discussed in the next sections of this chapter.

Several *adaptation strategies* consider that the biometric reference ref_j is composed of several biometric samples/templates (see III.3.1). This set of samples/templates is sometimes referred to as a *gallery* [Giot et al., 2012c, Biggio et al., 2012, Rattani et al., 2008a]. In line with this concept, adaptation can be defined as the addition and removal of samples/templates from a gallery.

The adaptation process can be performed either *online* or *offline* [Poh et al., 2012] (see III.3.4). In the online adaptation, the process is executed after each query sample is recognized by the biometric system. Basically, the adaptation process is triggered every time that the test/recognition is performed. In the offline adaptation, however, instead of triggering the adaptation process after each query recognition, the system waits to store a batch of biometric samples in the set $\mathcal{A}$ before adapting the biometric reference.

In this manuscript, the behavior of adaptation is determined by the *adaptation strategy*, which has an *adaptation criterion* (see III.3.2) and an *adaptation mechanism* (see III.3.5). As a conclusion, an adaptive biometric system is composed of a *classification algorithm* and an *adaptation strategy*.

The following table III.1 presents the recurrent terminology used within this manuscript.

III.3 Strategies to adapt the Biometric Reference

A number of adaptation strategies have been proposed in the literature. This chapter presents the most comprehensive collection of adaptation strategies that we are aware of. No previous review [Poh et al., 2009b, Seeger & Bours, 2011, Didaci et al., 2014, Poh et al., 2012] in the field of adaptive biometric systems have provided such an extensive and complete overview of the field.

This section presents adaptation strategies found in the literature, based on five distinctive aspects, as shown in Figure III.1(a):

- *Reference modeling*: How the biometric reference is modeled;

Table III.1 – Recurrent symbols of the manuscript

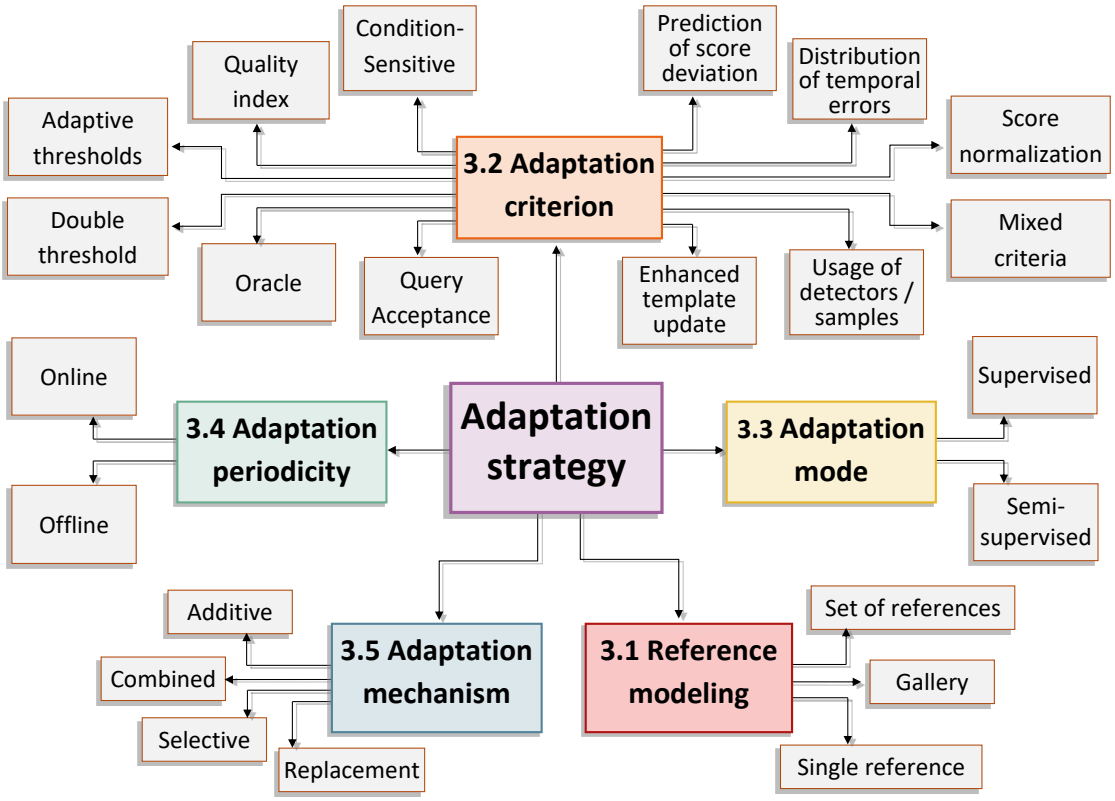
Terminology	Meaning/ Signification
$\mathcal{J}$	Set of user indexes registered in the biometric system
$j \in \mathcal{J}$	A registered user's index
ref_j	Biometric reference of a user j
$\mathcal{R} = \{ref_j \mid j \in \mathcal{J}\}$	The set of biometric references stored in the biometric system (biometric database)
q	A biometric query sample
$label^p$	Predicted label for the biometric query: *genuine* or *impostor*
θ_{verify}	Set of parameters for the test/recognition process (verification mode)
$\theta_{identify}$	Set of parameters for the test/recognition process (identification mode)
$\mathcal{A}$	Set of samples for the adaptation process
θ_{adapt}	Set of parameters for the adaptation process

- *Adaptation criterion*: The criterion chosen to perform adaptation or not;
- *Adaptation mode*: The method employed to assign the labels: supervised or semi-supervised;
- *Adaptation periodicity*: The periodicity in which the adaptation process is applied: online or offline;
- *Adaptation mechanism*: How the adaptation is performed (when the adaptation is satisfied).

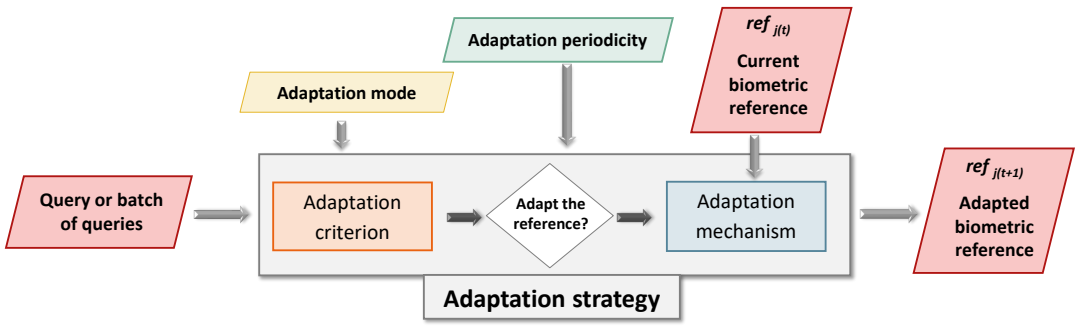
The five aspects illustrated by Figure III.1(b) are further discussed in the next subsections.

III.3.1 Reference Modeling

Biometric reference modeling strategies can impact the way it is adapted/updated over time. For instance, a speech signal can be represented as Mel-frequency cepstral coefficient (MFCC) features, whose density is modeled using a Gaussian Mixture Model. Thus, the



(a) Map of adaptive biometric systems aspects.



(b) Generic work-flow diagram of biometric adaptation process.

Figure III.1 – Overview of an adaptive biometric system.

resulting reference is a statistical model [Reynolds & Rose, 1995] storing the biometric features. Another example is when the k-nearest neighbor (k-NN) [Bishop, 2006] algorithm is used by a biometric system. In this case, its reference is a set of samples. A related concept is adopted when the biometric is a set of detectors, as in [Pisani et al., 2015b]. Each type of reference modeling may need distinct adaptation mechanisms.

Overall, three main categories of biometric references can be found in previous studies on adaptive biometric systems.

III.3.1.1 References containing a single sample/template

The biometric reference [Grabham & White, 2008] can be one good quality capture acquired at enrollment phase. Although this category has been used for physical modalities, it may not be reliable for behavioral biometrics. This is because a single sample is unlikely to capture enough variability usually present in behavioral modalities.

III.3.1.2 References built from several samples/templates

Several samples are acquired during the enrollment phase and stored in a *gallery*. In some studies, each sample is known as a *detector* [Pisani et al., 2015b]. Using galleries in adaptative biometric systems is a very common approach as it will be shown in the next sections.

III.3.1.3 Set of references

Several references per user are organized to represent different aspects of the biometric data [Lumini & Nanni, 2006]. Other examples are the biometric references used in [Giot et al., 2012c] and [Pisani et al., 2016], which contain two sub-references to support recognition and adaptation.

III.3.2 Adaptation Criterion

The adaptation criterion determines if the adaptation should be performed or not. Several criteria have been proposed in the literature:

- *Call for an oracle*: The decision to use a query for adaptation is taken by an oracle; it can be a human operator [Sukthankar & Stockton, 2001, Vandana, 2007] (supervisor/administrator).
- *Query acceptance*: Each accepted query is used to adapt the reference [Wang et al., 2012, Kang et al., 2007].

- *Double threshold*: In addition to the *decision threshold* already present for the recognition process, an additional *adaptation threshold* is adopted. Query samples that meet the *adaptation threshold* are used for adaptation. The *adaptation threshold* is usually more stringent than the decision threshold [Rattani, 2010]. Consequently, only highly confident queries are used for adaptation.

- *Quality index*: A quality index is used as part of the criterion to decide whether a given query is used for adaptation or not [Noval & López, 2008, Poh et al., 2010b].

- *Condition-sensitive*: It performs adaptation if conditions which are not already present in the biometric reference, are collected during operation, e.g., pose and illumination [Pagano et al., 2015].

- *Prediction of score deviation*: This criterion analyses the scores of the biometric system to estimate when the biometric reference should be adapted [Carls, 2009].

- *Distribution of temporal errors*: In an operational scenario, false non-matches can bring useful information to the system [Serwadda et al., 2013]. For example, a continuous sequence of false non-matches could mean that the reference has aged and the changes in the biometric features should be tracked.

- *Mixed criteria*: It is used under a multi-modal biometric system. For example, if the system uses a modality with high intra-class variability and another modality less affected by intra-class variations [Roli et al., 2007], adaptation can be performed if the number of non-matches by the first modality reaches a given threshold, while the second modality accepts the user. It can also occur at the feature level [Kekre & Bharadi, 2009] by using invariant features to confirm the adaptation of the variant features.

- *Enhanced template update (ETU)*: A system can be designed to model an individual with two sub-references: one genuine reference modeling the biometric features of the target individual, and one impostor reference, for the features of everyone else. Hence, each user is represented by two references. They can be used in different ways to support verification and adaptation. Both references are adapted over time. The genuine reference is adapted using queries accepted as genuine, while the impostor reference is adapted using the rejected query samples [Pisani et al., 2016].

- *Usage of detectors/samples*: It is based on the concept of checking the usage of biometric samples, referred to as detectors, in the biometric reference for matching.

The general principle is to discard unused detectors over time. Some variations were proposed [Pisani et al., 2015b, Pisani et al., 2014, Pisani et al., 2015a]. In Usage Control/Usage Control R/Usage Control 2 (see Section III.3.5), adaptation occurs if some detectors have not been recently used. Usage Control S additionally checks if at least two detectors match the input query.

- *Score normalization*: As discussed in Section III.2.2, some implementations output a score from the comparison of query sample to the biometric reference. Hence, based on this score, a threshold is applied to output the label (genuine or impostor) and to decide whether adaptation should occur. Score normalization [Poh et al., 2009a] refines the output score and, consequently, allows a better choice of thresholds. A preliminary study on the use of score normalization for supervised adaptation to handle different acquisition conditions is shown [Poh et al., 2010b]. Later, the use of score normalization in adaptive biometric systems was further studied in [Pisani et al., 2017], considering a biometric data stream context.

The first criterion requires an oracle to tell when adaptation should be performed [Sukthankar & Stockton, 2001] and it is not always feasible. Query acceptance is a simple criterion that avoids this problem [Kang et al., 2007]. However, it is prone to allow the inclusion of wrongly classified query samples into the genuine biometric reference. An alternative to deal with this problem is the double threshold [Rattani, 2010], which uses an additional threshold for adaptation. Nevertheless, the double threshold criterion usually only captures little variability, since only query samples with a high probability of the genuine user trigger the adaptation process. Although these methods can decrease the risk of wrongly including impostor samples in the genuine biometric reference, the expected performance gain thanks to the adaptation strategy is likely to be limited. Quality-index may also be used to only add high quality data to the biometric reference [Noval & López, 2008, Poh et al., 2010b].

The condition-sensitive criterion provides a way to avoid including redundant information into the biometric reference. Because it only adds new samples if new conditions are identified during the operation of the biometric system [Pagano et al., 2015].

It can also be possible to predict when adaptation should be performed by checking the score deviation [Carls, 2009]. However, the prediction may not be accurate if the biometric features from the users start to change in a different way over time. The distribution of errors over time also may indicate the need to adapt the biometric reference [Serwadda et al., 2013]. Nevertheless, this approach assumes that false non-matches can be detected reliably over time. For example, in border control, customers who have a refused entry would have to go

Table III.2 – Comparison of adaptation criteria

Criterion	Advantages	Drawbacks
Call for an oracle [Sukthankar & Stockton, 2001]	- The method is secure; - Uses only close genuine biometric samples from the biometric reference.	- There is no extra cost; - It is manual.
Query acceptance [Wang et al., 2012]	- The method is simple and allows automatic adaptation.	- Can include characteristics of wrongly accepted impostors in the genuine biometric reference.
Double threshold [Rattani, 2010]	- Can reduce the inclusion of impostors samples in the biometric reference by an additional (more stringent) adaptation threshold.	- Is only able to capture little variability.
Quality Index [Noval & López, 2008, Poh et al., 2010b]	- Avoids the use of low quality samples in the adaptation; - Can replace low quality data acquired in the enrollment procedure.	- Need to define the quality index, which can be modality dependent.
Condition-sensitive [Pagano et al., 2015]	- Excludes redundant information and can potentially reduce the size of the reference, saving computer resources.	- Sensitive to the the initial samples in the reference as well as the updating threshold.
Prediction of score deviation [Carls, 2009]	- Prediction of the moment to update the biometric reference.	- If the pattern in which the biometric features change over time, the prediction may not be accurate.
Distribution of temporal errors [Serwadda et al., 2013]	- Monitors the actual error to mitigate it.	- Requires a way to measure false non-matches over time.
Mixed criteria [Roli et al., 2007]	- Uses additional information from multiple biometric modalities.	- Requires more than one biometric modality, increasing costs.
Enhanced template update [Pisani et al., 2016]	- Combines a genuine and an impostor gallery to support both test and adaptation	- Classification errors may poison both galleries
Usage of detectors/samples [Pisani et al., 2015b, Pisani et al., 2014, Pisani et al., 2015a]	- Keeps the biometric reference updated by the patterns most frequently and recently present in the queries.	- May remove true user patterns from the biometric reference if they are not frequently present in the queries.
Score normalization [Poh et al., 2010b, Pisani et al., 2017]	- Refine the output score for a better threshold choice.	- Require additional data to normalize scores (a development or a cohort database depending on the normalization procedure).

to a separate queue for manual identity verification. Therefore, closely monitoring the error over time constitutes a viable criterion for adaptation.

Using multiple sources to support the adaptation criterion is observed in the mixed criteria [Roli et al., 2007] and the enhance template update [Pisani et al., 2016]. The former works with multiple biometric modalities in a multi-modal system, while the latter stores a genuine and an impostor model to support the decision to whether or not perform adaptation.

The usage of detectors for matching can also provide information to decide whether adaptation should be started. Various ways of using this information have been proposed [Pisani et al., 2015b, Pisani et al., 2014, Pisani et al., 2015a].

Score normalization is an alternative to refine the output score in adaptive biometric systems [Poh et al., 2009a, Poh et al., 2010b]. As a result, a better threshold choice can be done, improving the performance of the adaptation criterion. A previous work has applied score normalization to several adaptation strategies in a biometric data stream context [Pisani et al., 2017]. Applying score normalization requires additional data, either a development or a cohort database depending on the normalization procedure.

As discussed in this section, there are several criteria that can be adopted to decide whether adaptation should be performed or not. They rely on different aspects, such as score, quality, errors and usage. However, they are still prone to adversarial attacks, which could introduce impostor patterns into the genuine biometric reference [Biggio et al., 2015, Biggio et al., 2012]. To summarize the discussion so far on adaptation criteria, Table III.2 highlights their advantages and drawbacks.

III.3.3 Adaptation mode

Query samples are usually unlabeled. In some cases, however, the true label is received some time after the biometric system has classified them, similarly to data stream mining applications [Žliobaitė et al., 2015]. When query samples are unlabeled, semi-supervised adaptation is performed. When the data are labeled, supervised adaptation techniques can be used. This section briefly describes both techniques as two adaptation modes for adaptive biometric systems: supervised and semi-supervised adaptation.

III.3.3.1 Supervised adaptation

It uses true labels of the query samples for adaptation that are provided by an oracle, also known as an operator in this context. It has been extensively studied in the literature [Freni et al., 2008a, Freni et al., 2008b, Giot et al., 2011b, Uludag et al., 2004] as it is an easier approach, particularly when compared to its semi-supervised counterpart.

The samples can be obtained from different ways. For example, several enrollment sessions can be applied to each user [Uludag et al., 2004]. The newly acquired samples at each enrollment session are labeled and can be used to adapt the biometric reference. Of course, this approach can be time-consuming and expensive as it requires individuals to participate in several enrollment sessions. Moreover, an operator must supervise these enrollments to avoid errors. Another way to obtain labels for supervised adaptation is by manually labeling the captured data when the authentication system is in use, e.g., in an operator-assisted face recognition system [Sukthankar & Stockton, 2001]. However, this approach is not applicable to many contexts of application, especially when the operator or supervisor is not available.

III.3.3.2 Semi-supervised adaptation

It is a more realistic scenario [Jiang & Ser, 2002, Ryu et al., 2006, Poh et al., 2015b, Pisani et al., 2016, Bours & Ellingsen, 2018], where the labels are provided automatically by the biometric system. Note that in this case, however, the obtained labels can be wrong. It is semi-supervised as the system has the labelled samples from the initial enrollment and the unlabelled query samples. The concept is to automatically label the query samples in order to use them during the adaptation. There are two main ways to perform it: *self-training* [Rattani et al., 2011] or *co-training* [Blum & Mitchell, 1998]. Self-training is related to mono-modal authentication systems and uses samples classified by the same classifier to retrain it, while co-training is related to multi-modal authentication systems and consists of using the knowledge of one modality to help to label the other one.

To summarize, semi-supervised methods automatically provide labels to the collected queries, thanks to the classifiers, whereas supervised ones rely on an oracle. The main drawback with the semi-supervised approach is that these labels can be wrong in case of recognition errors.

In this manuscript, we are only interested in semi-supervised adaptation as it is a more realistic and complex scenario. An immediate question is: how often should the update be? This is presented next.

III.3.4 Adaptation periodicity

The adaptation process does not have to be performed each time a query satisfies an adaptation criterion (see Figures III.2 III.3). There are two main settings: delayed/offline adaptation and real time/online adaptation. We detail each of them in the following sections.

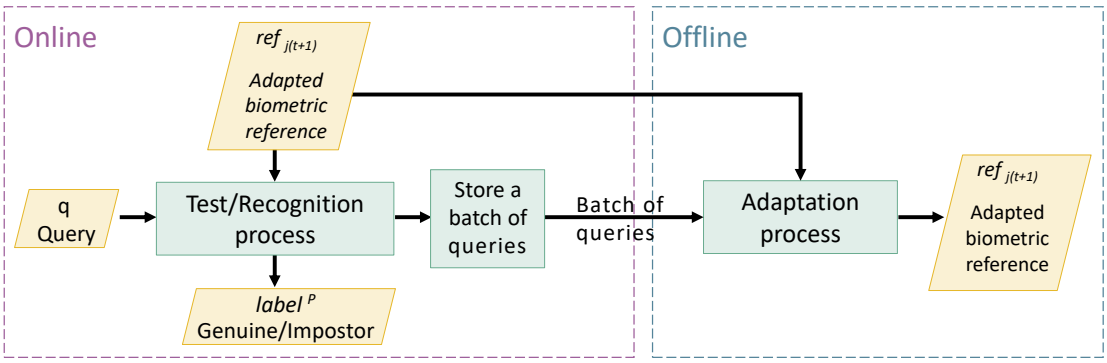


Figure III.2 – Delayed/Offline adaptation.

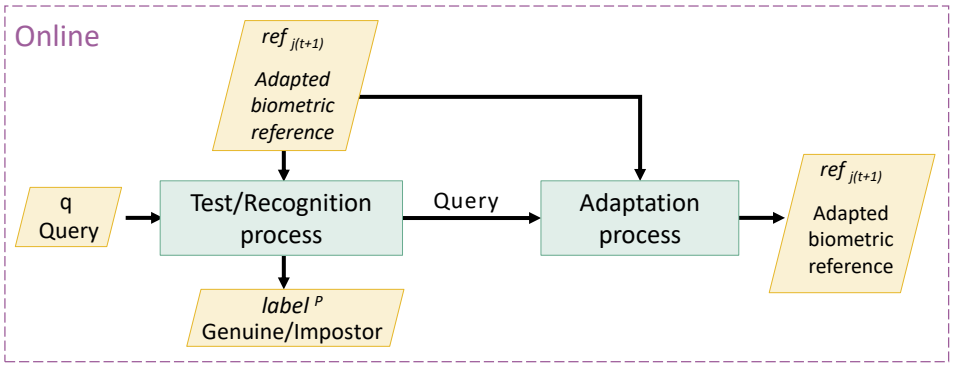


Figure III.3 – Real-time/Online adaptation.

Figure III.4 – Offline/Delayed vs Online/Real-time adaptation. The figure is simplified to correspond to the case where the adaptation criterion takes the decision just after the recognition process.

III.3.4.1 Offline/delayed adaptation

When the process is performed offline, queries are collected and stored in a buffer, before processing them as whole in a batch procedure. It is a common approach in the literature. However, the choice of the adaptation frequency remains an open issue. Which is the best strategy to adopt: waiting until enough samples have been collected or waiting for the expiration of a specific delay? As discussed later in Section III.3.6, the periodicity have been determined by the dataset division of sessions in previous studies. However, it is still an open question in practical application scenarios.

III.3.4.2 Online/real-time adaptation

This setting systematically performs adaptation after the decision criterion is met (often taken after the acceptance of the query [Giot et al., 2011b, Pisani et al., 2015a]). As the process is iteratively done, query per query, it mainly fits the semi-supervised adaptation

mode, where the adaptation system uses the label computed by the verification method on the selected query.

Offline adaptation has the advantage of a minimum performance impact during the recognition process, since no adaptation is done while the system is operating. The adaptation process can then be triggered when the system is not in use. Online adaptation, however, adds more processing time during the recognition process, as both recognition and adaptation processes are performed jointly. Nevertheless, it must also be noted that the online adaptation does not need to store a buffer, hence, it consumes less memory than the offline setting.

The adaptation periodicity can also affect the choice of the adaptation mechanism. Some mechanisms were designed for offline adaptation, such as Graph min-cut (see Section III.3.5) that needs a buffer of queries to build a graph, as part of its adaptation process.

III.3.5 Adaptation mechanism

As discussed before, an adaptation strategy is composed of various modules, and the last one to deal with is the adaptation mechanism, which finally adapts the biometric reference. All adaptation mechanisms presented in this section are suitable for references that are composed by a set of templates/samples, sometimes named as gallery [Giot et al., 2012c, Biggio et al., 2012, Rattani et al., 2008a]. Hence, the adaptation mechanism basically adds samples to the gallery and/or removes samples/templates from it. The reference is re-computed once the gallery is adapted. Four categories of adaptation mechanisms are presented here:

- *Additive mechanisms* receive a set of samples and add all (or some) of them to the gallery;
- *Replacement mechanisms* receive a set of samples and add all (or some) of them to the gallery but also remove some samples from the gallery;
- *Multi-gallery mechanisms* manage two (or more) galleries and can also apply distinct adaptation mechanisms to each gallery;
- *Selection mechanisms* select the most important samples in gallery to keep in order to avoid the gallery to indefinitely increase its size over time.

The above adaptation mechanisms are presented in the next sub-sections, discussing their advantages and drawbacks.

III.3.5.1 Additive mechanisms

An additive mechanism is based on the concept of progressively adding new patterns to the biometric reference. This mechanism can encode a higher variability of the user data, which can consequently avoid false non-match due to genuine intra-class variability.

One of the first attempts in this direction was proposed by Uludag *et al.* [Uludag et al., 2004]. The proposed mechanism, called *augment-update*, adds a set of new samples to the gallery of the user. Their experiments consider this new set of samples to be genuine. Subsequent works on the additive mechanism considered using the predicted labels instead of the true labels for adaptation. Some of them are described in the following:

• *Self-Update*: as described in [Rattani et al., 2013b], it is an implementation of self-training [Rattani et al., 2011] for adaptive biometric systems [Roli & Marcialis, 2006]. It has been extensively studied in the literature [Roli & Marcialis, 2006, Roli et al., 2008, Rattani, 2010, Giot et al., 2012a, Giot et al., 2011a, Akhtar et al., 2014]. The general concept is to add query samples classified as genuine to the gallery as depicted in algorithm 1. Usually, only those samples that meet a genuine similarity score above a given *adaptation threshold* are added to the gallery. Hence, Self-Update is commonly implemented together with *double threshold*, as described in Section III.3.2.

Algorithm 1: Self-Update [Roli & Marcialis, 2006] adaptation strategy for user j. We consider the implementation described in [Rattani et al., 2013c].

Input : $ref_{j(t)}, \mathcal{A}, \theta_j^{adapt} = \{adaptationThreshold\}$
Output : $ref_{j(t+1)}$

1 *All samples with similarity score similarityScore above the adaptationThreshold are used to adapt the gallery.*
2 $\mathcal{A}' \leftarrow \{a_i \in \mathcal{A} \mid similarityScore(ref_{j(t)}, a_i) > adaptationThreshold\}$
3 $\mathcal{G}(ref_{j(t+1)}) \leftarrow \mathcal{G}(ref_{j(t)}) \cup \mathcal{A}'$

Another related adaptation mechanism is the *Growing window* [Kang et al., 2007]. Growing window works similarly to Self-Update, however, it does not use the additional *adaptation threshold*. It can also be understood that it assumes that both decision and adaptation thresholds are the same. As a result, all queries classified as genuine are added to the gallery.

Concerning the adaptation periodicity, in the literature, Self-Update is frequently applied in a scenario of offline adaptation, where a batch of queries is received for adaptation from time to time. Conversely, Growing window is usually applied in online adaptation scenarios, where the adaptation process is executed after each query is processed.

• *Graph min-cut for template update*: it is an adaptation mechanism proposed by [Rattani et al., 2008a, Rattani et al., 2013a], which uses the *max-flow/graph min-cut* algorithm [Blum & Chawla, 2001]. This adaptation mechanism, detailed in algorithm 2, receives a batch of query samples and joins them to the current gallery of the user. Based on this data, a graph is generated, where each node represents a sample and each weighted link is a similarity score between samples. The graph-min cut divides the graph into two parts: source (genuine samples) and sink (impostors samples). The source represents the new gallery. The way that graph is generated implies that no sample in the initial gallery is removed during adaptation (all samples from the gallery are assigned infinite weight to the source/genuine node), justifying the categorization as an additive mechanism.

Algorithm 2: Graph min-cut adaptation strategy [Rattani et al., 2008a, Rattani et al., 2013a] for user j.

Input : $ref_{j_{(t)}}, \mathcal{A}, \theta_j^{adapt} = \{k\}$
Output : $ref_{j_{(t+1)}}$

1 $\mathcal{A}' \leftarrow \mathcal{G}(ref_{j_{(t)}}) \cup \mathcal{A}$
2 $graph \leftarrow buildGraph(\mathcal{A}', k)$
3 $(\mathcal{S}^{source}, \mathcal{S}^{sink}) \leftarrow applyGraphMinCut(graph)$
4 *By definition,* $\mathcal{G}(ref_{j_{(t)}})$ *is a subset of* $\mathcal{S}^{source}$.
5 $\mathcal{G}(ref_{j_{(t)}}) \leftarrow \mathcal{S}^{source}$

• *Adaptation using harmonic function*: a work from [Rattani et al., 2012] proposes an adaptation mechanism using harmonic functions, which makes use of probabilistic semi-supervised learning introduced in [Zhu et al., 2003]. Similarly to the previous adaptation mechanism based on graph min-cut, this adaptation mechanism also receives a batch of query samples and joins them to the current gallery of the user. The joined set of samples is used to compute an adjacency matrix, which is then applied to obtain an harmonic function for the set of query samples. The obtained harmonic function is employed to determine which queries are added to the gallery.

Self-update refers to a category of adaptation mechanisms that uses only one classifier [Roli & Marcialis, 2006, Rattani et al., 2013b]. It is vulnerable to the mistaken introduction of impostor samples in the gallery. Although, this problem is faced by most adaptation mechanisms, its impact is worst in the case of additive ones, since the gallery keeps growing and none sample is removed.

Of course, this could be avoided by a very high *adaptation threshold*. Nevertheless, this also means that only those genuine queries very close to the current reference would be

accepted for adaptation. Since they are already close to the reference, they could not bring enough new information and larger changes would not be captured. This illustrates that the configuration of the adaptation threshold deeply impacts the performance of the adaptation mechanism.

Considering the graph-based mechanism, its authors claim they can capture larger intra-class variabilities than Self-Update [Rattani et al., 2013a]. However, this mechanism as well as the one based on harmonic functions [Rattani et al., 2012] needs more computer resources than Self-Update, and the computations can become intensive, particularly if the gallery and the set of queries are large.

Commonly, additive mechanisms have the drawback of indefinitely increasing the size of the gallery, which could lead to problems in terms of memory usage. A possible way to mitigate it would be to use selection mechanisms described in Section III.3.5.4 after the additive mechanism is executed. Table III.3 summarizes the advantages and drawbacks discussed here.

Table III.3 – Comparison of additive mechanisms.

Mechanism	Advantages	Drawbacks
Self-Update [Roli & Marcialis, 2006]	- Simple to implement.	- The adaptation threshold can be difficult to define: low values may imply in several impostor samples included in the gallery while high values can prevent proper adaptation to genuine data.
Graph min-cut for template update [Rattani et al., 2008a, Rattani et al., 2013a]	- Able to capture higher intra-class variability than Self-Update.	- The computations can become intensive.
Adaptation using Harmonic function [Rattani et al., 2012]	- Obtain good performance even with few labeled samples.	- The computations can become intensive.

III.3.5.2 Replacement mechanisms

This mechanism adds new samples to the gallery over time, similarly to an additive one. However, it is also able to remove the samples to avoid the problem of indefinitely increasing the gallery size.

Again, one of the first attempts following this concept is from Uludag *et al.* [Uludag et al., 2004], which presented the *batch-update*. This mechanism receives a set of samples and uses it as the new gallery. Thus, the entire previous gallery is discarded. In their experiments, it was considered that the true label is provided for the new set of samples, which may not

be a feasible assumption in practice. Several replacement mechanisms are presented in this section. All of them assume that the query or the set of query samples received as input are classified as genuine.

• *Sliding/Moving window*: this mechanism was described in [Kang et al., 2007], though it can also be found under the name of First In First Out (FIFO) [Freni et al., 2008a, Scheidat et al., 2007]. This mechanism receives a set of query samples (the set can contain just one sample) and adds them to the gallery by removing the same amount of oldest samples, thus keeping the gallery size constant over time as depicted in algorithm 3. Double threshold criterion can be used with this mechanism. As a consequence, only samples that obtain a similarity score above a given *adaptation threshold* is added to the gallery. Another related adaptation mechanism is adopted in [Grabham & White, 2008], which works similarly to growing window until the gallery reaches a maximum size, when it uses a sliding window for the adaptation.

Algorithm 3: Sliding/Moving window [Kang et al., 2007] adaptation strategy for user j.

Input : $ref_{j_{(t)}}, \mathcal{A} = \{\mathbf{q}\}, \theta_j^{adapt} = \{label^p\}$
Output : $ref_{j_{(t+1)}}$

if $label^p = genuine$ **then**
  $g' \leftarrow oldest(\mathcal{G}(ref_{j_{(t)}}))$
  $\mathcal{G}(ref_{j_{(t+1)}}) \leftarrow \mathcal{G}(ref_{j_{(t)}}) - \{g'\}$
  $\mathcal{G}\left(ref_{j_{(t+1)}}\right) \leftarrow \mathcal{G}\left(ref_{j_{(t+1)}}\right) \cup \{\mathbf{q}\}$
end

• *Replacement based on MDIST and DEND*: Freni *et al.* [Freni et al., 2008a] proposed replacement mechanisms based on the operating principle of MDIST and DEND clustering algorithms [Uludag et al., 2004]. The general concept is to add a new query sample and remove another one from the gallery, thus keeping the same gallery size after the adaptation. For such, all possible gallery variations are evaluated (each time a different sample is removed). The scores among all samples are computed for each gallery variation. This process is also performed for the unmodified gallery. Then, the average score for each gallery is obtained. Based on this average score, the gallery is chosen according to one of the two strategies here: for MDIST, the gallery which has the maximum average score is chosen, whereas, for DEND, the chosen reference is the one corresponding to minimal average score.

• *Least frequently used (LFU)*: LFU was presented in [Scheidat et al., 2007, Freni et al., 2008a] and consists in adding the received query sample to the gallery and removing the

least frequently used ones. It is mandatory to maintain the number of times each sample of the gallery is used to authenticate the user.

• *Least recently used (LRU)*: LRU proposes to replace the least recently used sample of the gallery by the new query sample [Scheidat et al., 2007]. A first method is to use a time stamp for each sample of the gallery, but it can be too expensive. The authors then suggest to use the *clock algorithm*, a special case of the second-chance approach [Scheidat et al., 2007].

• *Extended replacement*: it computes a relevance attribute for each sample of the gallery based on its usage for matching and performs replacement based on this attribute [Scheidat et al., 2007]. The sample with lowest value for this relevance attribute is removed and the new query sample is added to the gallery.

• *Usage Control*: it is based on the concept of checking the usage of detectors (biometric samples) in the biometric reference for matching to perform adaptation. The more recently (and frequently) used detectors are kept in the biometric reference, while the remaining detectors are removed. Four versions are proposed: Usage Control, Usage Control R, Usage Control S and Usage Control 2 [Pisani et al., 2015b, Pisani et al., 2014, Pisani et al., 2015a]. If the adaptation criterion for them is met, a detector (or a set of detectors) is removed from the biometric reference. In Usage Control/Usage Control R, the mechanism first selects those detectors less recently used. Among them, the less frequently used is removed. Usage Control S works similarly, but it has a more stringent adaptation criterion (at least two detectors should match the input query). Usage Control 2 can remove more than one detector since it removes all detectors not recently used. As as example, the algorithm for Usage Control/Usage Control R is shown in Procedure 4.

• *Transfer learning-based*: in [Çeker & Upadhyaya, 2016, Çeker & Upadhyaya, 2017], the authors presented adaptive mechanisms based on transfer learning to update SVM classifiers [Taylor & Stone, 2009]. Given an SVM trained on the enrollment data, the adaptive mechanism is capable of adapting it using later acquired labeled samples. Note that these mechanisms do not use a gallery as the other ones presented here. However, as it replaces the older user model with the newer adapted using transfer learning, it is classified as a replacement mechanism in this manuscript.

As stated in the beginning of this section, a key advantage of replacement mechanism is that it can avoid increasing the gallery indefinitely over time. The crux of maintaining the gallery size is that when a new sample is added to the gallery, another one has to be removed. This section presented several ways to choose which samples are replaced. The simplest one is sliding window/FIFO [Kang et al., 2007, Scheidat et al., 2007], which simply replaces the oldest sample(s). This mechanism assumes that the most recent samples are more

Algorithm 4: Usage control R [Pisani et al., 2015a] adaptation strategy for user j.

Input : $ref_{j_{(t)}}, \mathcal{A} = \{\mathbf{q}\}, \theta_j^{adapt} = \{label^p, MAX\_RU\}$
Output : $ref_{j_{(t+1)}}$

if $label^p = genuine$ **then**
  (Checks the detectors from the newest to the oldest one.) **for** $i \leftarrow 1$ **to** $length(\mathcal{G}(ref_{j_{(t)}}))$ **do**
    if d_i *matches* $\boldsymbol{q}$ **then**
      $usage_C(d_i) \leftarrow usage_C(d_i) + 1$ $usage_R(d_i) \leftarrow MAX\_RU$ $Y \leftarrow \mathcal{G}(ref_{j_{(t)}}) - \{d_i\}$ *The attributes ($usage_C$, $usage_R$) of the first detector that matches the query are updated (usually $MAX\_RU = 10$).*
      for $k \leftarrow 1$ **to** $length(Y)$ **do**
        if $usage_R(d_k) > 0$ **then**
          $usage_R(d_k) \leftarrow usage_R(d_k) - 1$
        end
      end
      break
    end
  end
  $L \leftarrow \{d_i \in \mathcal{G}(ref_{j_{(t)}}) \mid usage_R(d_k) \leq 0\}$
  if $L \neq \emptyset$ **then**
    $L' \leftarrow \{ order(d_i \in L)\ by\ usage_C(d_i)\}$ *All the detectors in Y are ordered by $usage_C$.* $\mathcal{G}(ref_{j_{(t+1)}}) \leftarrow \mathcal{G}(ref_{j_{(t)}}) - \{l'_1\}$ *The detector with the lowest $usage_C$ is removed* $\mathcal{G}(ref_{j_{(t+1)}}) \leftarrow \mathcal{G}(ref_{j_{(t+1)}}) \cup \{\mathbf{q}\}$ *and the accepted query is added to the biometric reference.*
    else
      $ref_{j_{(t+1)}} \leftarrow ref_{j_{(t)}}$ *Note that no detector is added nor removed (just the parameters $usage_C$ and $usage_R$ are updated).*
    end
  end
end

representative, though it may not be always the case. Table III.4 summarizes the discussion for the replacement mechanisms.

MDIST and DEND [Freni et al., 2008a] can be computationally intensive if the gallery is large, since it requires to compute the scores among all samples for several gallery variations. Since MDIST maintains the gallery with the highest average score, the obtained gallery has less variability among the samples than the gallery obtained by DEND (which keeps the gallery with lowest average score). The authors mention that MDIST is based on the idea of

Table III.4 – Comparison of replacement adaptation mechanisms.

Mechanism	Advantages	Drawbacks
Sliding window [Kang et al., 2007]	- Simple, just replaces the oldest samples considered less representative.	- Oldest samples may be more representative.
MDIST and DEND [Freni et al., 2008a]	- MDIST can exploit common representative characteristics, while DEND is able to represent larger intra-class variability.	- Both can be computationally intensive.
Least frequently used (LFU) [Scheidat et al., 2007]	- Replace less frequently used patterns.	- May not replace a frequent used sample that becomes unrepresentative.
Least recently used (LRU) [Scheidat et al., 2007]	- Replace less recently used patterns.	- May be expensive, since it requires to store when each sample is used.
Extended replacement [Scheidat et al., 2007]	- Assigns a relevance attribute to each sample, that can be used to replace less representative samples.	- Problem similar to LFU, since it may not replace a frequent used sample that becomes unrepresentative.
Usage Control [Pisani et al., 2015b, Pisani et al., 2014, Pisani et al., 2015a]	- Does not change the biometric reference if all patterns are being used, which could mean that the user characteristics have not changed.	- May not properly adapt the reference if all patterns were recently used and the user starts to change its characteristics.
Transfer learning [Çeker & Upadhyaya, 2016, Çeker & Upadhyaya, 2017]	- Can adapt SVM models without the need to retrain it.	- Uses labeled samples to adapt the SVM model.

keeping samples that are very similar to exploit common representative characteristics, while DEND is able to represent larger intra-class variability.

A technical report [Scheidat et al., 2007] presented three adaptation mechanisms that replace samples considering their usage, although none of them were experimentally evaluated. LFU replaces the most frequently used sample. If a sample is too frequently used for some time, it can be hard to replace it later if it becomes unrepresentative of the current user data. Moreover, older samples tend to be more used, making the mechanism subject to replace newer samples over time, which may not be the most suitable choice. LRU then replaces the least recently used, but it may be expensive to run the mechanism since it needs to know when each sample is used. Extended replacement then assigns a relevance attribute to each sample and replaces the ones with lowest values for this new attribute. Nevertheless, this mechanism is subject to a problem similar to LFU, since a too much used sample which becomes unrepresentative will not be easily replaced.

Usage Control keeps only those detectors more frequently and recently used. It can overcome some of the issues of the previous algorithms based on usage of samples as discussed in [Pisani et al., 2015b]. For example, even if a detector/sample is used too many times and becomes unrepresentative, it could be quickly replaced if it is not used for a while. Hence, even if the frequency of usage of a detector/sample is the highest among all samples, it can be replaced if it has not been used recently. Usage Control 2 [Pisani et al., 2015a] implements another interesting proposal which is the possibility of having a gallery of variable size. Some versions of Usage Control does not always replace a sample [Pisani et al., 2015b]. If it considers that the current biometric reference is representative, the replacement does not occur, as described in the criterion Usage of detectors/samples in Section III.3.2.

Recent works use transfer learning to adapt SVM models [Çeker & Upadhyaya, 2016]. The proposal obtained good results. However, the evaluation methodology described in these works mention that the samples used for adaptation are labeled. Nevertheless, in a practical scenario, the true labels may not be available. It is still unclear whether it can obtain good performance if predicted (and not true) labels are used for adaptation.

III.3.5.3 Multi-gallery mechanisms

A multi-gallery mechanism manages two or more galleries/models to perform adaptation and can apply different adaptation mechanisms to each one. This can be interesting to combine the benefits of different adaptation mechanisms into a single one. Some implementations are presented next:

• *Double parallel*: it consists of using two galleries, where one is adapted by Growing window and the other is adapted by Sliding window [Giot et al., 2012c] as depicted in algorithm 5. The classification and adaptation then considers the average of the scores obtained by both galleries. An incremental version of this mechanism, designed for the case when the classification algorithm of [Magalhães et al., 2005] is used was presented in [Pisani et al., 2015a]. This new version allows to use the growing window without the unlimited memory issue.

• *Co-Update*: it is an implementation of the concepts from Co-training [Blum & Mitchell, 1998] to adaptive biometric systems [Roli et al., 2007, Rattani et al., 2008b]. We consider the implementation described in [Rattani et al., 2013c]. Co-Update, as described in algorithm 6, is applied to a multi-modality scenario, with two galleries, one for each biometric modality (e.g. one for face and another for fingerprint). It assumes that two biometric samples (one for each modality) are provided for each query. Then, if the classifier trained for modality A confidently classifies the corresponding query, the query for modality B is added to the corresponding gallery. The opposite also applies, if the classifier trained for modality B

Algorithm 5: Double parallel [Giot et al., 2012c] adaptation strategy for user j.

Input : $ref_{j_{(t)}}, \mathcal{A} = \{\mathbf{q}\}, \theta_j^{adapt} = \{adaptationThreshold\}$
Output : $ref_{j_{(t+1)}}$

This strategy keeps two models: T^1 *and* T^2.
$score1 \leftarrow similarityScore(T^1(ref_{j_{(t)}}), \mathbf{q})$
$score2 \leftarrow similarityScore(T^2(ref_{j_{(t)}}), \mathbf{q})$
$fusionScore \leftarrow (score1 + score2)/2$
if $fusionScore > adaptationThreshold$ **then**
    This strategy manages two galleries: $\mathcal{G}^1$ *and* $\mathcal{G}^2$. *After the adaptation of the galleries, the models* T^1 *and* T^2 *are recomputed.*
    $\mathcal{G}^1(ref_{j_{(t+1)}}) \leftarrow adaptUsingGrowing(\mathcal{G}^1(ref_{j_{(t)}}, \mathbf{q})$
    $\mathcal{G}^2(ref_{j_{(t+1)}}) \leftarrow adaptUsingSliding(\mathcal{G}^2(ref_{j_{(t)}}, \mathbf{q})$
end

confidently classifies the query for its modality, the query for modality A is added to the gallery of modality A. Co-Update is similar to the cross-training mechanism presented in [Poh et al., 2014, Poh et al., 2015b]. Another application of Co-training to adaptive biometric systems was presented in [Zhao et al., 2011], where a single modality was considered (face recognition). In their work, each of the two classifiers considered a different view of the face image.

Poh *et al.* [Poh et al., 2014] also discuss the application of Co-training to adaptive biometric systems. These works studied a system where there is one gallery for face recognition and another for speech recognition. Taking advantage of the availability of two modalities, logistic regression combines the face and speech scores to obtain the final fused score which is then used to infer the samples for adaptation. The proposed strategy was named fusion-based co-training.

Algorithm 6: Co-Update adaptation strategy for user j. This chapter considers the implementation described in [Rattani et al., 2013c].

Input : $ref_{j_{(t)}}, \mathcal{A}, \theta_j^{adapt} = \{adaptationThreshold^1, adaptationThreshold^2\}$
Output : $ref_{j_{(t+1)}}$

$\mathcal{A}'^1 \leftarrow \{a_i^1 \in \mathcal{A}^1 \mid similarityScore(ref_{j_{(t)}}^2, a_i^2) > adaptationThreshold^2\}$
$\mathcal{A}'^2 \leftarrow \{a_i^2 \in \mathcal{A}^2 \mid similarityScore(ref_{j_{(t)}}^1, a_i^1) > adaptationThreshold^1\}$
$\mathcal{G}^1(ref_{j_{(t+1)}}) \leftarrow \mathcal{G}^1(ref_{j_{(t)}}) \cup \mathcal{A}'^1$
$\mathcal{G}^2(ref_{j_{(t+1)}}) \leftarrow \mathcal{G}^2(ref_{j_{(t)}}) \cup \mathcal{A}'^2$

• *Enhanced template update (ETU)*: it was recently proposed to make use of all queries, including those classified as impostor [Pisani et al., 2016], whereas adaptation mechanisms are generally only interested in the queries classified as genuine. In order to implement it, ETU manages two galleries, one for queries classified as genuine and another for queries classified as impostor. The ETU framework then employs both galleries to support classification and adaptation.

• *Ensembles*: El Gayar *et al.* [El Gayar et al., 2006] proposed to use several classifiers in an ensemble configuration to address the problem of having a limited amount of labelled enrollment samples. Another work which also applied ensembles for adaptive biometric system is [Pisani et al., 2015c], where different adaptation mechanisms were combined in an ensemble.

One of the first multi-gallery mechanisms proposed in the literature is Co-Update [Roli et al., 2007, Rattani et al., 2008b, Rattani et al., 2013c], applied to multi-modal systems. This adaptation mechanism can adapt the biometric reference to larger changes due to the use of two biometric modalities. For example, in case of an abrupt change in one biometric modality, while the other does not change, the biometric system would be able to capture this large change and adapt the reference. Otherwise, an adaptation mechanism that uses just one gallery would not be able to decide whether this abrupt change is an impostor attempt or not.

Later, Double parallel [Giot et al., 2012c] was proposed. It manages two galleries for a single modality, each adapted by a different adaptation mechanism. One gallery uses Growing, thus preserving the initial user patterns, while the other gallery uses Sliding, thus maintaining only the most recent user patterns. As a result, Double parallel can combine the models obtained from both galleries to support classification and adaptation. Since Double Parallel uses Growing, one of its galleries can increase without any limit over time. In [Pisani et al., 2015a], the authors proposed an incremental solution to deal with this problem for the classification algorithm of [Magalhães et al., 2005].

Most adaptation mechanisms only consider galleries for genuine data and, as a consequence, they discard queries classified as impostor. Enhanced template update (ETU) [Pisani et al., 2016], conversely, manages a genuine and an impostor gallery. Hence, all queries, even those classified as impostor, are used for adaptation. ETU then combines both galleries to support classification and adaptation.

Ensembles of classifiers have also been used in the literature of adaptive biometric systems [El Gayar et al., 2006, Pisani et al., 2015c]. Although the use of additional classifiers can result in higher use of computer resources, the robustness of the classification and adaptation can be increased. The fusion-based co-training proposed by Poh et al [Poh

et al., 2014] can be considered as an example of this approach as well, where a classifier is associated with a biometric modality and both results are fused.

A summary of the discussion is shown in Table III.5.

Table III.5 – Comparison of multi-gallery adaptation mechanisms.

Mechanism	Advantages	Drawbacks
Co-Update [Roli et al., 2007, Rattani et al., 2008b, Rattani et al., 2013c]	- Can adapt the reference even for large intra-class variation.	- Requires two biometric modalities working in parallel with aging patterns not correlated.
Double parallel [Giot et al., 2012c]	- Can combine two adaptation strategies, one preserving initial patterns (Growing) and another maintaining only the latest patterns (Sliding).	- Can increase the amount of used memory indefinitely, although a solution for a specific classification algorithm has been presented in [Pisani et al., 2015a].
Enhanced template update [Pisani et al., 2016]	- Manages a genuine and an impostor gallery, making use of all received queries to adapt them.	- Classification errors can result in unreliable information on both galleries.
Ensembles [Pisani et al., 2015c]	- Increased classification reliability by the use of ensembles.	- Needs more processing time than single classifier system due to the use of several of them in the ensemble configuration.

III.3.5.4 Selection mechanisms

Selection mechanisms, also named template selection [Freni et al., 2008b], are used to select representative samples/templates for the user. These mechanisms can be used to reduce the size of the user gallery after adaptation [Uludag et al., 2004]. Some implementations are presented next:

• *Selection based on clustering* [Uludag et al., 2004]: it is based on the algorithms used for replacement shown in Section III.3.5.2. DEND applies a hierarchical clustering algorithm, which outputs a dendrogram on which a pre-defined number of clusters is identified. For each cluster, the medoid element (sample) is kept in the user gallery, while the other samples are discarded. The other mechanism, MDIST, sorts the samples by their average distance to all other samples. Those samples with the lowest average distance are kept in the user gallery, while the others are discarded. For both mechanisms, the number of samples to be kept needs to be defined beforehand. This number should be lower than the amount of samples in the gallery.

• *Selection based on editing*: in [Freni et al., 2008b], the authors proposed the use of algorithms based on nearest neighbor algorithm to select the most representative samples for a user. The following algorithms were used: Condensed NN (CNN) [Hart, 1968], Selective NN (SNN) [Ritter et al., 1975], Reduced NN (RNN) [Gate, 1972] and Edited NN (ENN) [Wilson, 1972].

Both methods, selection based on clustering [Uludag et al., 2004] and based on editing [Freni et al., 2008b], can be used to reduce the gallery size after adaptation. This can be particularly important for additive mechanisms, such as Self-Update [Roli & Marcialis, 2006, Rattani et al., 2013b]. Freni *et al.* [Freni et al., 2008b] compared both types of mechanisms and showed that editing mechanisms can obtain better performance than clustering mechanisms. A summary of this discussion in shown in Table III.6.

Table III.6 – Comparison of selection mechanisms.

Mechanism	Advantages	Drawbacks
Selection based on clustering	- Can reduce the size of the gallery using clustering algorithms.	- Can be computationally intensive for large galleries.
Selection based on editing	- Can reduce the size of the gallery using NN-based algorithms.	- When strong gallery size limitations are imposed, the output gallery can be negatively impacted. - Can be computationally intensive for large galleries.

III.3.6 Evaluation methodology

Sadly, there is no standard methodology to evaluate adaptive biometric systems in the literature [Giot et al., 2012d]. A number of methodologies that differ in several aspects have been adopted, as discussed in the next sections.

III.3.6.1 Impostor samples in the adaptation process

Recent studies dealing with adaptation consider that the set of biometric samples for adaptation $\mathcal{A}$ (Equation (III.4)) is a set of samples without the true label. Thus, only labels obtained from the classification algorithm are available, so they are subject to wrong prediction from the classifier. This better simulates a practical scenario where true labels are usually not available. Consequently, the set $\mathcal{A}$ may contain impostor samples resulted from misclassification.

However, early investigation on adaptive biometric systems did not consider the possibility of impostor attack during the adaptation. According to [Poh et al., 2012, Rattani et al., 2013c], impostor attacks during the adaptation process were not considered in [Roli & Marcialis, 2006, Roli et al., 2007]. Another study that did not considered impostor samples in the set $\mathcal{A}$ is [Kang et al., 2007]. As mentioned in [Giot et al., 2011a, Giot et al., 2012b], the experiments in [Kang et al., 2007] only employed true genuine samples for adaptation. In [Kang et al., 2007], each user was enrolled using 10 samples and, for test, there were 75 genuine samples plus 75 impostor samples. Although not entirely clear, the graphs from Figure 4 of that paper indicate that a separate set of genuine samples was used for adaptation.

III.3.6.2 Ratio of impostor samples

A related aspect is the ratio of impostor samples that can be part of the adaptation set $\mathcal{A}$. A high ratio can result in several errors during the adaptation process. In [Rattani et al., 2013c], the adaptation set $\mathcal{A}$ contains 10 genuine samples and 5 random impostor samples, so the ratio of impostors is 33.3%. Another study [Giot et al., 2012c] adopted the ratio of 30% of impostor samples. It assumes a scenario where the genuine user is the most frequent user of the biometric system, which is a valid assumption in general.

Later, different ratios of impostor samples were investigated in [Giot et al., 2013] where the samples from the first session are used for enrollment. Then, the samples from the remaining sessions are used for test and adaptation using *pools*. A *pool* is defined as a sequence of query samples, containing both genuine and impostor. The ratio of impostors in the *pools* ranged from 30% to 80%. One pool was generated for each session in the dataset. By doing this, the performance metrics could be assessed over time, one for each session.

The same ratio of impostors of 30% was also adopted in [Pisani et al., 2015a, Pisani et al., 2016, Pisani et al., 2017]. However, in these studies, a distinct method was adopted to select the impostor samples. The evaluation methodology adopted there, named *user cross-validation for biometric data streams*, divides the list of user indexes using cross-validation, so k folds are obtained (each fold is a disjoint sub-set of the user indexes). One fold is regarded as the unregistered set of users and the remaining folds form the registered set of users $\mathcal{J}$. The experiments are executed for all k combinations of folds, so all users are considered once as an unregistered user. Among the 30% of impostor samples, there is 50% probability of obtaining an impostor sample from the unregistered set (external attack simulation) and 50% probability of obtaining a sample from another user $i \neq j$ (internal attack simulation) as impostor.

III.3.6.3 Adaptation to time vs condition

Template aging is one of the main motivations for adapting a biometric reference. This is clear in keystroke dynamics in which the typing rhythm changes over time. However, the biometric reference may need adaptation to deal with different acquisition conditions too, which is not necessarily due to aging. For instance, in face recognition, if the enrollment uses samples of just one pose (e.g. frontal), the system would need to adapt the reference later to include variations in the pose of the same user.

The methodology described in the last section from [Giot et al., 2013] is one that mainly deals with adaptation due to aging. This is because the first session is used for training and next ones are left for test and adaptation, following the chronological order.

Conversely, the experiments in [Poh et al., 2014] is an example of methodology which mainly deals with adaptation to different conditions. That work used a dataset which has data under three different conditions: controlled (sessions 1-4), degraded (sessions 5-8) and adverse (sessions 9-12) [Bailly-Bailliére et al., 2003]. Session 1 was used for enrollment, then sessions 2 to 4 for test. Next, session 5 was used for adaptation and sessions 6 to 8 for test. Finally, session 9 was used for adaptation and sessions 10 to 12 for test. Impostor samples were included in the adaptation sets too.

Adaptive biometric systems can be used to adapt the biometric reference to changes either due to time or due to different capture conditions. Some studies on physical biometric modalities seem to mainly deal with changing conditions instead of changes uniquely due to time, which is the case of that study.

III.3.6.4 Poisoning attacks to adaptation

Poisoning attacks in adaptive biometric systems consist of progressively introducing impostor samples in the adaptation process, in a way that the biometric reference is modified until it can better recognize an impostor. As a result, it may also not be able to recognize the actual genuine user anymore. Such attacks are not simulated in most evaluation methodologies for adaptive biometric systems.

The work that claims to be the first to raise such issue in the area is [Biggio et al., 2012]. In order to evaluate this attack, the authors used a dataset for face recognition containing 60 samples per user. A random subset of 10 samples was used for the enrollment and another subset of 10 samples was used for parameter tuning. The remaining 40 samples were then used for the test. Then, a separate set of poisoning samples was used to adapt the biometric reference. Nevertheless, their work only considered that the biometric reference is adapted with impostor patterns from the generated poisoning set. This may not correspond to a

practical scenario, since both genuine and impostor samples can be used for adaptation and, consequently, the negative effect of poisoning could be reduced.

III.3.6.5 Separate and joint sets for test/adaptation

Most previous evaluation methodologies can be divided into two groups: *separate sets* or *joint set* for test and adaptation. A previous review in the area adopted this criteria to classify performance assessment approaches [Poh et al., 2009b]. In the *separate sets* approach, the adaptation and test sets are disjoint and, consequently, samples used for adaptation are not part of the test. This approach assumes that the biometric system can stay a period only adapting the biometric references (without performing test/recognition). Later, the adapted biometric reference is fixed to perform recognition only. Some recent studies have also adopted the *separate sets* approach [Biggio et al., 2012, Poh et al., 2014]. However, this approach may not be the best choice in some cases since it does not make an optimal usage of the available data. This is due to the non-overlapped adaptation and test sets. The optimal usage of the dataset is a critical issue in the area, particularly in view of the limited amount of large datasets for studying adaptive biometric systems.

The *joint sets* for test and adaptation approach, on the other hand, share data for test and adaptation, so both sets are not disjoint. This approach also better represents a practical scenario, where the system, once deployed, has to perform the recognition of all query samples and use this data for adaptation. Hence, the system does not stop the recognition for a period of adaptation.

An important work in the area which proposed an evaluation methodology following the *joint sets* approach is [Rattani et al., 2013c]. A similar methodology was used in another work from the same authors in [Rattani et al., 2013a]. Their methodology was based on the DIEE dataset, which has several sessions per user, each containing 10 samples. The following steps are performed:

- Part A (enrollment): the first 2 samples of the first session ($t = 1$) are used for enrollment.
- Part B (adaptation): for each user, an adaptation set $\mathcal{A}$ is formed by the samples from current session t plus five random impostor samples. The first session used for adaptation is $t = 1$, however, in this particular case, the first two samples are discarded since they were already used for enrollment, while, in the other sessions, all 10 samples are part of the adaptation set. This adaptation set is then presented to the adaptation strategy to perform adaptation.

- Part C (test): the adapted biometric reference is used to test on the next session. The first test session is $t + 1$. Biometric samples from the same session from all other users are regarded as impostors to compute the performance metrics. Note that the biometric reference is not adapted during the test. When the test is finished on session $t + 1$, Part B is launched again, though on session $t + 1$ this time. The adapted biometric reference is then tested on session $t + 2$ and so on.

Note that in this methodology, the last session is used only for test and the very first session is only part of the enrollment and adaptation. However, all other sessions are used for both adaptation and test, meaning that it mainly adopts the *joint sets* for test and adaptation approach. As a result, the amount of samples used for both adaptation and test is increased.

Another evaluation methodology that follows the *joint sets* approach is [Giot et al., 2012c] and its modification to include variable impostor ratios too [Giot et al., 2013]. As described earlier in this chapter, a *pool* is generated for each session. The *pools* are used for test and adaptation, so the same data is used by both processes. The work from [Pisani et al., 2015a, Pisani et al., 2016] also adopted this approach as the same biometric data stream used for test is also the input for the adaptation process.

III.3.6.6 Online vs Offline adaptation

As discussed in Section III.3, the periodicity of adaptation can change. There are two general categories: offline and online adaptation. In the offline adaptation, the biometric reference keeps unchanged for some time, then it is adapted at specific periods. Conversely, in the online adaptation, the biometric reference is adapted after each query is presented to the biometric system.

The methodology from [Rattani et al., 2013c] described earlier in this chapter is an example of offline adaptation, since the biometric reference keeps unchanged during the test, while the methodology adopted in [Giot et al., 2013] deals with online adaptation. The *pool* is presented query by query to the biometric system, which performs recognition and then adapt the biometric reference.

III.3.6.7 Chronological order

Usually, the evaluation of adaptive biometric systems respects the chronological order of the biometric samples. In order to properly evaluate how the biometric system adapts the biometric reference to changes over time, the enrollment should be done using the oldest samples and the test using the newest samples, in chronological order. As a result, the biometric reference is adapted to progressive changes observed over time. Modifying the

order of the samples during the test can change how the biometric reference is adapted and, therefore, if the goal is to study changes due to time, the obtained results would be unreliable.

In [Giot et al., 2013], for instance, the samples from the *pool* are randomly interleaved (between genuine and impostors samples), but the chronological order of the genuine samples is maintained. Another work that respects the chronological order is [Mhenni et al., 2016], but is not the case for all studies in the area, such as [Biggio et al., 2012], which studied the effect of poisoning attacks. In that study, random samples were used for enrollment, so test samples may be older than the enrollment ones. Note that is does not mean the methodology adopted by [Biggio et al., 2012] is wrong, as their purpose was not to evaluate adaptation to genuine samples, but, instead, study poisoning attacks.

III.3.6.8 Division into sessions and biometric data streams

As discussed in the previous section, several evaluation methodologies used the division into sessions to guide the assessment of the biometric systems. In [Rattani et al., 2013c, Rattani et al., 2013a], for example, the session division information is used to guide when the adaptation process is launched. In other studies, such as [Giot et al., 2012c, Giot et al., 2013], the session division information is used to guide the generation of the *pools*, as one *pool* is obtained from each session. The decision threshold also may change over sessions since the results are reported in terms of EER.

Such a kind of information regarding the session division may not be available in a practical scenario. In light of this fact, in the studies in [Pisani et al., 2015a, Pisani et al., 2016, Pisani et al., 2017] which used the *user cross-validation for biometric data streams* methodology, a biometric data stream is generated for each user ignoring the session division. It works by joining all the sessions into a single one, then the first samples are used for enrollment (where parameter tuning is performed too) and the remaining samples are used to form a biometric data stream. This biometric data stream is a sequence of queries presented, sample by sample, to the biometric system, which will return the $label^p$ for each query. The decision to adapt or not the biometric reference in the meantime is up to the adaptation strategy. As a result, the decision to adapt a biometric sample is taken by the adaptive biometric system without the help of additional information, such as the session division.

III.4 Conclusion

The aim of this chapter has been to provide a wide review of adaptive biometric systems, covering aspects such as formalization, terminology, sources or variations that motivates the use of adaptation, adaptation strategies, evaluation methodology and open challenges. To the

best of our knowledge, this is most up-to-date and complete review of adaptive biometric systems.

Thanks to the developed taxonomy for adaptation strategies presented in this chapter, the reader is able to have a broad view of works in adaptive biometric systems and easily compare them. Adaptation strategies were divided into components, namely: reference modeling, adaptation criterion, adaptation mode, adaptation periodicity and adaptation mechanism.

Another contribution of this chapter is discussing the distinct evaluation methodologies that have been adopted in previous work. The way to evaluate an adaptive biometric system differs on the way to evaluate standard biometric authentication systems. Common evaluation metrics have been redefined to be properly expressed in the context of adaptive systems and specific metrics have been presented as well.

Regarding this detailed state of the art, we will present the proposed contributions in order to remedy the reported problems.

CHAPTER IV

Single Enrollment for Keystroke Dynamics with Adaptive Template Update

IV.1 Introduction

Cyber-attacks have spread all over the world to steal information such as trade secrets, intellectual property and banking data. Facing the danger of the insecurity of saved data (personal, professional, official, etc), keystroke dynamics was proposed as an interesting, non-intrusive, inexpensive, permanent and weakly constrained solution for users. Based on the typing rhythm of users, it improves logical access security. Nevertheless, it was demonstrated that such an authentication mechanism would need a larger number of samples to enroll the typing characteristics of users. Moreover, these registered characteristics generally undergo aging effects after a time span. Different solutions have been suggested to remedy these variability problems, including template adaptation. In this chapter, we propose a double serial adaptation strategy that considers a single-capture-based enrollment process. When using the authentication system, the template of users and the decision/adaptation thresholds are updated. Experimental results on three public keystroke dynamics datasets show the benefits of the proposed method.

IV.2 Target objectives

In the literature, most studies have required more than twenty captures to create the reference template during the enrollment phase [Giot et al., 2011b], as depicted in Table IV.1. However, considering usability, it is not really operational to ask users to type their password 20 times.

Table IV.1 – Gallery size in enrollment phase for some systems in literature

Works	Gallery size in the enrollment phase
[Çeker & Upadhyaya, 2017]	20-40-60-80-100-120-140-160
[Ceker & Upadhyaya, 2016]	15
[Pisani et al., 2016]	40
[Yu & Cho, 2004]	50
[Obaidat & Sadoun, 1997]	112
[Killourhy et al., 2009]	200

In fact, it has been demonstrated that performances increase with the number of enrolled samples in the template. In contrast, another study [Giot et al., 2011b] used only five samples

per user. The authors considered that "5" is the maximal number of samples for usability reasons in industrial conditions. Indeed, even if the keystroke dynamics modality has proved its efficiency in several scientific research papers, it is still not fully adopted in industrialized applications, unlike other morphological modalities such as the fingerprint (*e.g.*, fingerprint scanner [Fernandez-Saavedra et al., 2016], Touch ID [Marasco & Ross, 2015], etc.) and the face (*e.g.*, video cameras on consumer devices [Smith et al., 2015], etc.). This is basically owing to the need of several typing captures during the enrollment phase to create the reference template that describes the typing rhythm of the users. It is not the case for real applications for which the password is usually requested only once, when creating an account. As shown in Table IV.1, for all the published research papers, the learning phase requires a large number of samples which generally exceeds 20 according to [Giot et al., 2011b].

Besides, the problem of the tedious enrollment phase, keystroke dynamics particularly suffers from large intra-class variation, as well as other behavioral modalities. Thus the need of a specific adaptive strategy. First of all, it serves to enrich the typing manner description by increasing the size of the reference. Second of all, it solves the problem of intra-class variation.

IV.3 Proposed adaptive strategy

We put forward a novel adaptive method that considers a limited number of samples used to create a user's reference while keeping a good performance. Indeed, the user introduces the password only once, when creating a new account. Thus, the reference is composed of a single sample. Afterwards, for each successful authentication, the reference is updated in a transparent way. Avoiding the enrollment phase, the growing window mechanism serves to increase the size of the reference to capture more intra-class variations. Once the size of the reference reaches 10 samples, the sliding window will be considered in order to limit the number of samples saved in the reference. Moreover, the process detailed in Figure IV.1, contains different contributions as follows :

- We consider a preprocessing step which intends to eliminate the noise in the captured characteristics (peaks corresponding to hesitation, disturbances or workload of the computer).
- We use a single sample to create a user's reference while avoiding the tedious step of typing the same password several times in the enrollment phase.
- We use a *GA-KNN verification method*: It is based on the optimized combination of multi-distance metrics for the KNN classifier, which shows a better performance. This

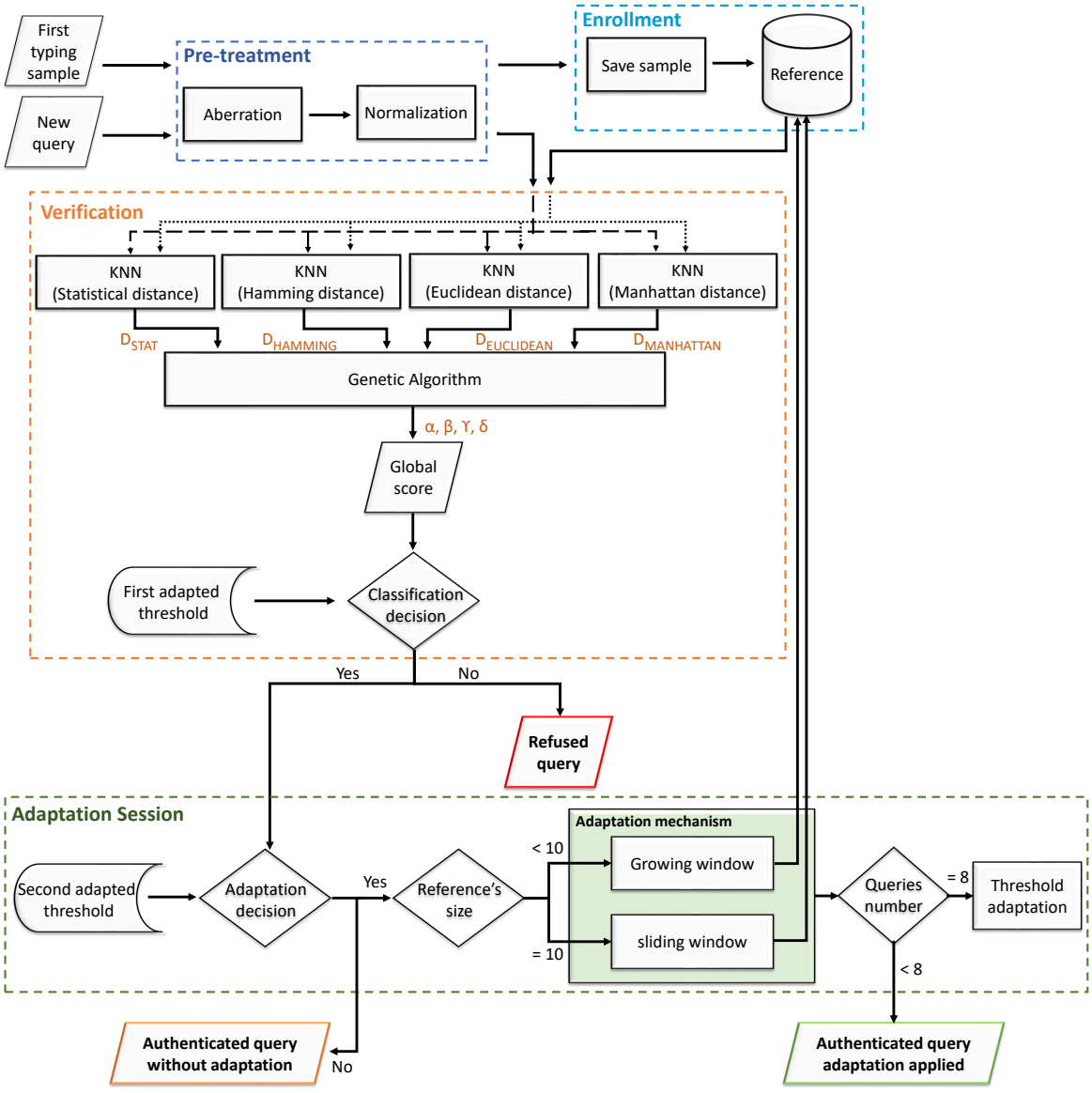


Figure IV.1 – Proposed method

combination is ensured by vote parameters that are optimized by GA and updated during the use of the system.

- We propose to adapt the reference and the used thresholds over time. Hence, our method also considers the decision of the *adapted thresholds* criterion (user and time-dependent).

- We resort to a *double serial mechanism*: This progressive adaptation mechanism combines the growing-window and sliding-window mechanisms (respectively before and after reaching the number of required samples, empirically set at 10).

Thus, a new authentication framework is proposed in addition to the adaptation strategy. Indeed, previous works use baseline authentication method to evaluate their update system. Now, we detail our contributions in each step of the process.

IV.3.1 Preprocessing phase

To describe the keystroke dynamics of one user, we are interested in temporal information extracted from digraph transition times:

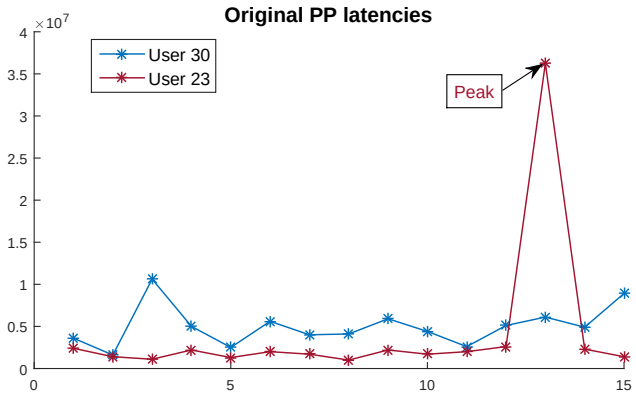
- PP: time difference between the press events of two successive keys ;

- RR: latency between the release events of two successive keys;

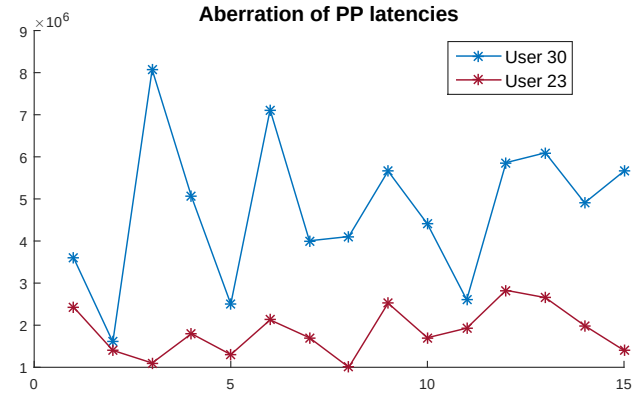
- PR: time duration between a one-key press event and its following key release event;

- RP: time duration between a one-key release event and its following key press event.

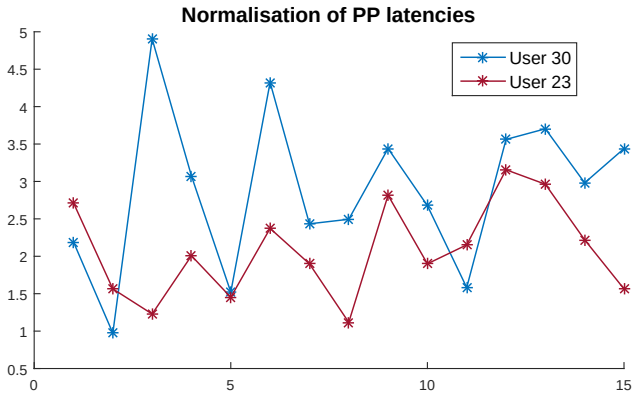
Hence, the characteristic vector C is composed of these temporal informations $C = [PP\ PR\ RR\ RP]$. These characteristics undergo preprocessing steps, as demonstrated in Figure IV.2. We first apply an aberration correction to the acquired characteristics aiming to detect the peaks where the user takes an abnormally longer time to type a password. In fact, these peaks do not describe a user's typing manner. They are generally caused by a disturbance, hesitation time, etc. For that purpose, we first define the peaks as the i^{th} characteristic value $C(i)$ three times greater than the i^{th} value of the standard deviation vector of the reference $\sigma_C(i)$. The peak is then replaced by the i^{th} value of the mean vector μ_C of the reference. This correction is applied to two peaks of the characteristic vector at most. Equation IV.1 summarizes this preprocessing step, in case it is applied to the characteristic vector C.



(a) Acquired characteristic vector of PP latencies



(b) Characteristic vector of PP latencies after applying aberration correction



(c) Characteristic vector of normalized PP latencies

Figure IV.2 – Successive preprocessing steps: (b) Aberration correction and (c) Normalization, applied to (a) Characteristic vector of PP latencies.

$$\begin{cases} IF\ (C(i) \geq (\ 3 \times \sigma_C(i))) \ THEN \\ C(i)\ =\ \mu_C(i) \end{cases} \tag{IV.1}$$

where:
i is the index of the i^{th} character of a vector ;
C is the characteristic vector of each keystroke dynamics acquisition presented to the preprocessing phase;
σ_C is the standard deviation vector of the reference. When the reference contains an only one sample, σ_C is a vector of fixed values (which are the standard deviation value of the one sample reference);
μ_C is the mean vector of the reference.

After that, data normalization is carried out by dividing the characteristic vector by the standard deviation σ of the reference (to ensure a standard deviation of these features to 1), as depicted in Equation (IV.2). In fact, the normalization is applied to reduce the order of the magnitude of latencies from 10^6-order to 10^0-order values. Thus, the allocated memory space is reduced as well as the execution time.

This normalization is applied when the gallery contains at least two samples. Actually, when the reference contains only one sample, it is divided by the standard deviation which is a value and not a vector. So, all elements of the reference vector are divided by the same value. Whereas when the size of the reference is larger than 1, each element of the reference is divided by the corresponding element of the standard deviation vector.

$$C(i)\ =\ \frac{C(i)}{\sigma_C(i)} \tag{IV.2}$$

By applying the aberration correction and normalization steps, the erroneous data are almost removed. Thus, we obtain a sample composed of four characteristic vectors containing the information necessary to model the users' keystroke dynamics.

IV.3.2 Enrollment phase

Several biometric authentication systems, essentially those based on face and fingerprint modalities [Ryu et al., 2006, Rattani et al., 2007], use a single sample in the enrollment step. This is not the case for keystroke dynamics systems, since they are based on a behavioral modality that quickly changes over time.

According to the literature, the minimal number of samples used during the enrollment phase to create the reference is 5 samples [Giot et al., 2011b]. In this contribution, we use characteristics extracted from only a single sample to create gallery $\mathcal{G}_j$ of user j, in the enrollment phase. Therefore, the proposed method fits the industrial and operational application conditions, for which a user introduces a password only once when creating an account.

Indeed, the number of samples the user must type during the enrollment phase is a constraint that can penalize the authentication system. Most papers from the literature mentioned in this work required at least 5 samples to build the reference template. This is not operational. Indeed, users will be quickly annoyed. Therefore, we have chosen to consider a single sample, to alleviate the enrollment phase while keeping satisfying performances, especially when the purpose is the security of an industrial application. Using a single sample during the enrollment phase:

- is easier to achieve;
- ensures a lower computation time;
- fits the industrial application conditions;
- meets usability requirements.

IV.3.3 Verification phase

This phase aims to decide whether to authorize or deny an access for a claimed user. We judge the K Nearest Neighbor (KNN) approach to be the most appropriate classifier as it has proved to be efficient for keystroke dynamics modalities [Akhtar et al., 2014, Pisani et al., 2016], hence the competitive performances. Indeed, a single sample in the reference will not be efficient for the training phase of the other classifiers like NN or SVM. Since the KNN classifier can be applied with a variety of distances, several distance metrics are tested.

IV.3.3.1 Distance metrics analysis

The classification is ensured with a KNN classifier. Knowing that the KNN classifier can be used with different distance metrics, we propose to evaluate its performances with different metrics for the purpose of choosing the best ones.

Choice of distances :

The process of the suggested method is detailed in Figure IV.3. It is the same general process but without the calculation of the weighted vote combining the scores obtained by the

chosen distances. This process is used to choose the distances that offer the best performance among those tested.

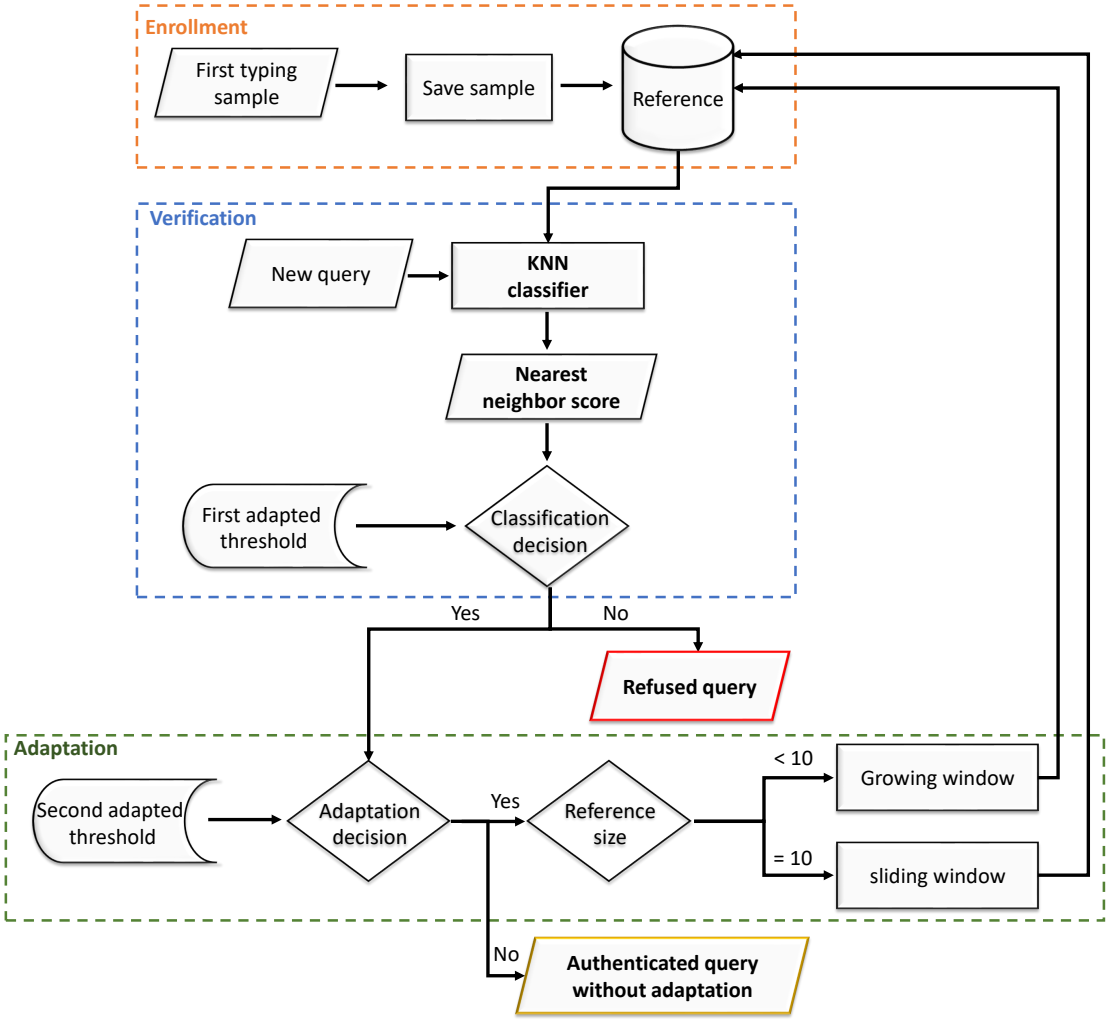


Figure IV.3 – Description of the keystroke authentication process

Regarding the obtained results, presented in the following, four main distance demonstrated better perfromances than the other tested ones: The Statistical [Hocquet et al., 2007], Hamming , Euclidean and Manhattan distances are considered to obtain four respective partial scores D_{STAT}, $D_{HAMMING}$, $D_{EUCLIDIAN}$ and $D_{MANHATTAN}$, as represented in Equation (IV.3). Thereby, each novel query is labeled by the KNN classifier using these four distances described below:

- Statistical distance: It is widely used for classifying keystroke dynamics data. Based on extracting statistical values from each biometric feature (mean and standard deviation), it has the advantage of being easy to calculate and offering competitive performances.

This distance is well known for its competitive performances and its calculation speed while being used for the keystroke dynamics authentication [Hocquet et al., 2007].

- Hamming distance: It consists in calculating the percentage of different coordinates between the novel query and the reference samples.
- Euclidean distance: It is a simple distance metric often used with a KNN classifier. It is defined as the square root of the sum of the squares of the differences between the corresponding coordinates of the new query and the reference samples.
- Manhattan distance: It computes the sum of the differences of the corresponding components of the new query and the reference samples.

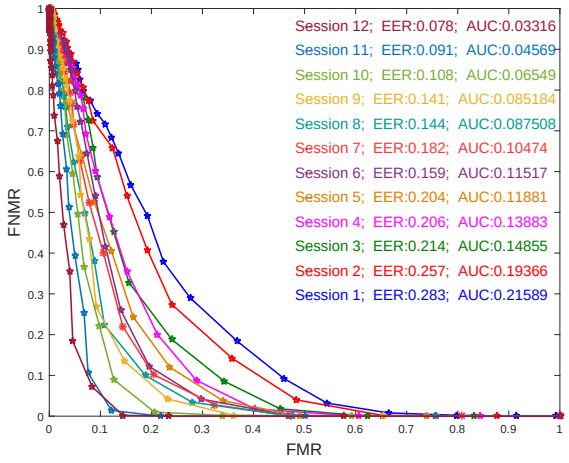
$$
\begin{aligned}
D_{STAT} &= 1 - \frac{1}{n}\sum_{i=1}^{n} e^{-\frac{|\, q_i - \mu_i \,|}{\sigma_i}} \\
D_{HAMMING} &= (\, \#\,(\, q_j \neq \mathcal{G}_j(i)\,)\,/\,n\,) \\
D_{EUCLIDIAN} &= \sqrt{\sum_{i=1}^{n} (\, q_j - \mathcal{G}_j(i)\,)^2} \\
D_{MANHATTAN} &= \sum_{i=1}^{n} |\, q_j - \mathcal{G}_j(i)\,|
\end{aligned}
\tag{IV.3}
$$

where:
q_j is the claimed query of the user j, $\mathcal{G}_j(k)$ is the k^{th} reference sample of the user j, m is the number of the samples in the reference $\mathcal{G}_j$, μ is the mean vector of the reference, and σ is the standard deviation vector of the user reference, and i varying from 1 to n where n is the length of the password.

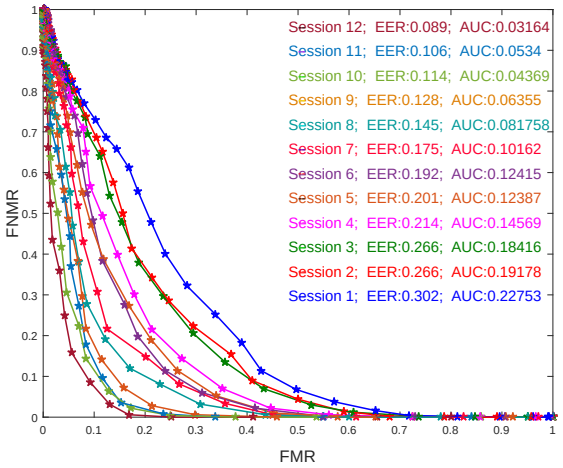
Distance assessment :

Although the reference initially contains only a single sample, the obtained results are promising. Figure IV.4 depicts the DET curves with the associated EER and AUC performances for the twelve adaptation sessions of the different experimentations applied to the GREYC 2009 database. Figure IV.5 illustrates the ROC curves and the performances (EER, AUC) of the first and the last adaptation sessions obtained using both databases.

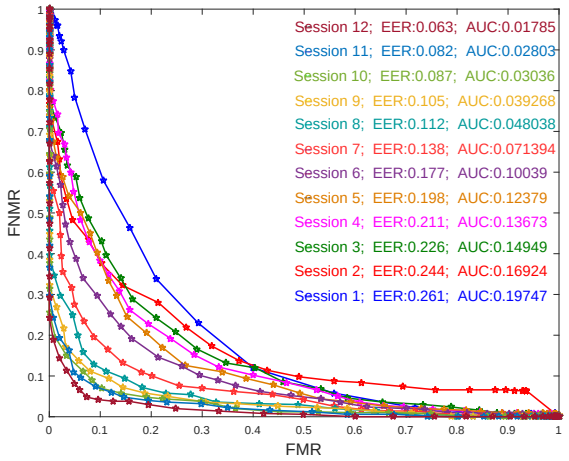
We choose four distance metrics to associate to the KNN classifier because we test a large number of distances, but those that demonstrate competitive performances are hamming, statistical, euclidean and Manhattan. Other tested metrics, were much less efficient than those considered as depicted in Table IV.2. Comparing the metrics with each other, we note that the



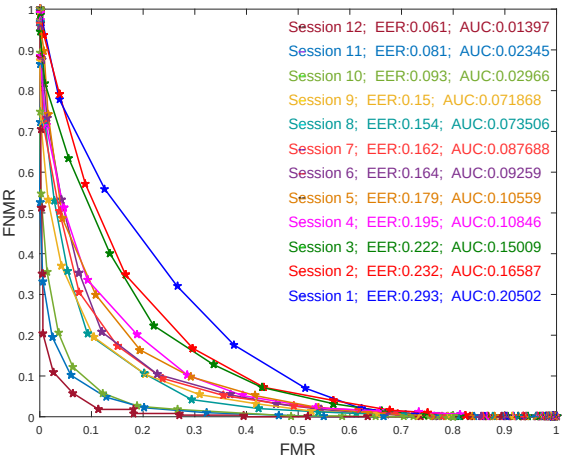
(a) Euclidean distance.



(b) Manhattan distance.



(c) Statistical distance.



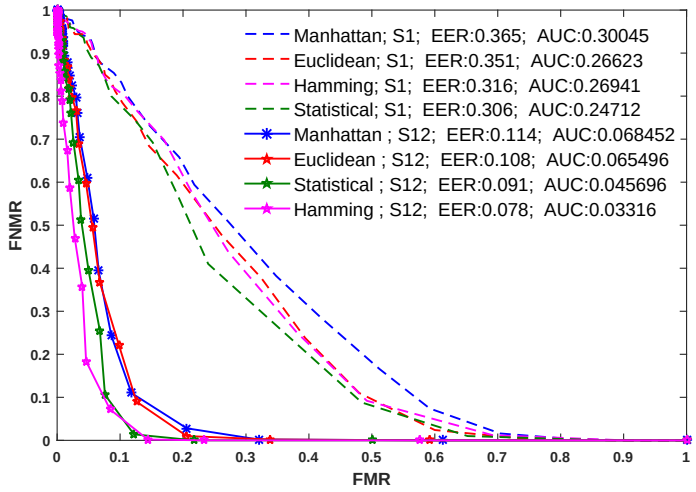
(d) Hamming distance.

Figure IV.4 – DET curves evolving over all adaptation sessions (GREYC 2009 database) and the associated performances (EER, AUC)

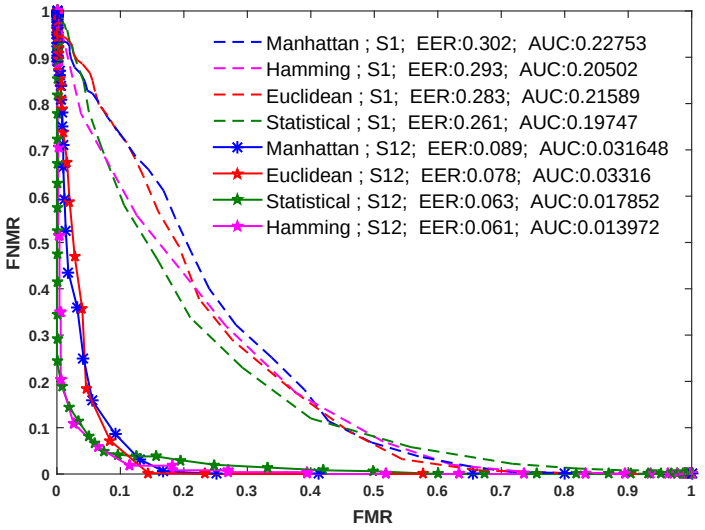
hamming distance and the statistical one perform better than others for the two considered databases.

We compare our approach with that of Giot *et al* [Giot et al., 2011b], in which the authors applied the average mechanism based on 3 different classifiers: Support Vector Machine (SVM), neural network and statistical distance. Thereby, an examination of the classifiers' performance is essential. Table IV.3 summarizes the compared results.

The best performance achieved in [Giot et al., 2011b] is an EER equal to 6.96%, while using an SVM classifier and the reference was composed of 5 samples as minimum size



(a) WEBGREYC database.



(b) GREYC 2009 database.

Figure IV.5 – DET curves of the first and the last adaptation session (S1,S12) and the associated performances (EER, AUC)

and 15 samples maximum. In the present study, we use the same database as in the work of [Giot et al., 2011b], thus obtaining two better performances: an EER equal to 6.3% using the KNN based on the statistical distance, and 6.1% using the KNN based on the hamming distance. We will benefit from the minimisation of the size of the reference while keeping better performances to facilitate the industrialisation of the keystroke dynamics modality. In addition, the KNN classifier compared to other classifiers, has the advantage of a low computing time which facilitates its deployment on the web server.

Table IV.2 – Comparison of performances obtained by different distances.

Distance	EER	AUC
City block distance	0.108	0.045
Chebychev distance	0.214	0.144
Minkowski distance	0.163	0.079
Correlation: One minus the sample linear correlation between observations (treated as sequences of values).	0.276	0.175
Jaccard: One minus the Jaccard coefficient, which is the percentage of nonzero coordinates that differ.	0.128	0.053
Spearman: One minus the sample Spearman's rank correlation between observations (treated as sequences of values).	0.196	0.119

Table IV.3 – Comparison of the chosen classifier with those of previous work for GREYC 2009 database.

Adaptive mechanism	Reference size Minimum	Reference size Maximum	Classifier	EER	AUC
Double	1	10	KNN (Hamming)	6.1%	0.013
serial	1	10	KNN (Statistical)	6.3%	0.017
mechanism	1	10	KNN (Euclidean)	7.8%	0.033
(Proposal)	1	10	KNN (Manhattan)	8.9%	0.031
Average	5	15	SVM	6.96%	-
mechanism	5	15	Neural network	8.75%	-
[Giot et al., 2011b]	5	15	Statistical	10.75%	-

IV.3.3.2 GA-KNN combination

To enhance the performances obtained by each metric distance apart, the KNN classifier is used with a multi-distance vote strategy assured by a Genetic Algorithm.

In Equation (IV.3), q_j is the query that claims to belong to user j. Hence, it is matched against its biometric reference $\mathcal{G}_j$. We use these four metrics because we have tested different distance metrics separately, and these ones have demonstrated better performances. The global score $Score_j$ is the weighted sum of the four partial scores. For each user j, we calculate the global score according to Equation (IV.4):

$$\begin{aligned} Score_j &= \alpha \times D_{STAT} + \beta \times D_{HAMMING} \\ &\quad + \gamma \times D_{EUCLIDIAN} + \delta \times D_{MANHATTAN} \end{aligned} \tag{IV.4}$$

where $\alpha, \beta, \gamma, \delta$ are the vote parameters. The calculation of these parameters will be further detailed in the "adaptation mechanism" point in the next section IV.3.4.

The calculated score is compared to a previously set verification threshold. As a result, it is very critical to define the appropriate threshold. Two types of thresholds have been defined in the literature:

- Global threshold: During the use of the system, a constant and unique threshold is fixed for all users.
- Individual thresholds: During the use of the system, user-specific thresholds are considered.

In [Giot et al., 2011b], the authors compared these two types of thresholds for the keystroke dynamics modality and showed that the best performances were obtained with the individual threshold. Thereby, we opt for this type of threshold in our experiments. In the next section, we show how to make both thresholds (decision and adaptation) time-dependent.

IV.3.4 Adaptation phase

This subsection presents an innovative adaptation method. It essentially updates both decision and adaption thresholds.

IV.3.4.1 Thresholds adaptation

For the proposed method, we use the double threshold criterion to make the adaptation decision [Rattani, 2010]. Two thresholds are used to make two successive decisions. The global score $Score_j$ is compared to a first threshold (decision one) to verify a user's identity. After acceptance, the same score is again compared to a second threshold to decide whether to use the query for adaptation. This adaptation criterion has been deeply used for adaptive systems concerning different modalities (face and fingerprint [Rattani, 2010], as well as keystroke dynamics [Giot et al., 2012c]).

The choice of the update threshold is very important. Actually, a strict threshold does not capture intra-class variability. On the other hand, a very high threshold raises the possibility of including imposter information to the gallery. In the literature, the decision threshold is

chosen using one of the following approaches: opting for the same threshold for all users; or utilizing a specified threshold for each user [Drygajlo et al., 2009]. It can be empirically or automatically defined, depending on the security level to reach.

It is known that the measured system's performance is different depending on the targeted choice [Hocquet et al., 2006] [Hosseinzadeh & Krishnan, 2008]. Moreover, it has been demonstrated that the individual threshold approach is more advantageous in terms of calculated error percentage [Giot, 2012]. Some studies [Drygajlo et al., 2009, Rattani et al., 2011] have analyzed the influence of age progression on the classifier thresholds and have proved that there is a conditional dependency between age progression and classifier scores. Thus these works have adapted the used thresholds and obtained better performances, but threshold adaptation has been mainly utilized for the face modality. These studies demonstrated that the variation in the users' characteristics over time would influence the scores obtained by the classifiers. Consequently, it is better to update the thresholds in order to cope with these variations.

In this work, we propose to adapt both thresholds and we demonstrated in [Mhenni et al., 2016] that updating the used thresholds would improve the system performance. An individual decision threshold T_j^{i+1} of session $(i+1)$ is adapted by decreasing it with a coefficient depending on the average of the mean vector of reference μ and the standard deviation of the standard deviation vector σ, as indicated in Equation (V.3). The initial thresholds are fixed to $EER \simeq 3\%$ (the best performance we have obtained).

$$T_j^{i+1} = T_j^i - e^{-\frac{\mu_j}{\sigma_j}} \tag{IV.5}$$

Thereby, the thresholds are specific to the user and to the session at the same time.

Highlighting of the adapted thresholds:

To highlight the advantages of the adaptation of the used thresholds we compared it to the other thresholds: global and individual. Thus, we applied the experimentation whose parameters are summarized in the table IV.4.

We firstly present the results obtained by applying this approach on the GREYC 2009 database. Figure IV.6 represents results validated with fixed, individual and variable thresholds. The basic scenario without any template update is "None". The scenarios using a template update strategy are "Sliding" and "Growing".

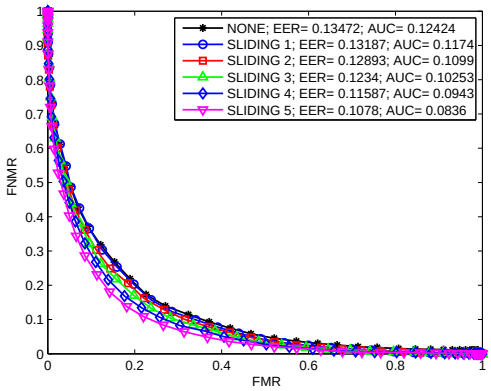
Table IV.4 – Experiment parameters to highlight the adapted thresholds

Parameter	Value
Modality	Keystroke dynamics
Authentication method	Statistical classifier
Update decision	Double threshold
Update mode	Semi-supervised
Update periodicity	Delayed
Update strategy	None, sliding window, growing window
Number of sessions	5 sessions
Respect to chronology	Yes
Enrollment samples	10 user's samples
Verification samples	10 user's samples, 10 impostor's samples
Evaluation metrics	EER , ROC

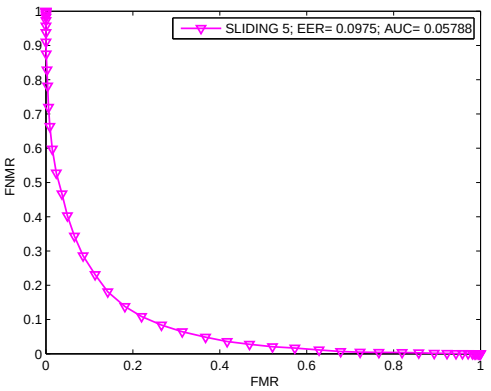
The same process is repeated to the same database, but with a growing window update strategy. Figure IV.7 illustrates the results we obtained. For the Web-GREYC database, we show the results in Figure IV.8.

Experimentally, taking into consideration the intra-class variation, the characteristics of the users change over time. Thus, by decreasing the threshold, we only pick the most similar characteristics to the reference. However, we do not suddenly reduce the threshold. Instead, we start with a high threshold (but which differentiates between samples of authentic user and those of impostors) in the first update session. Consequently, we introduce into the reference some new users' samples that are different. From one session to another we slightly decrease the threshold so that we can capture less dissimilarity. Finally, samples similar to the modified reference (containing new samples added in last sessions) are captured. This is explained by the fact that by mastering the password, there is a noticeable stability in the typing manner.

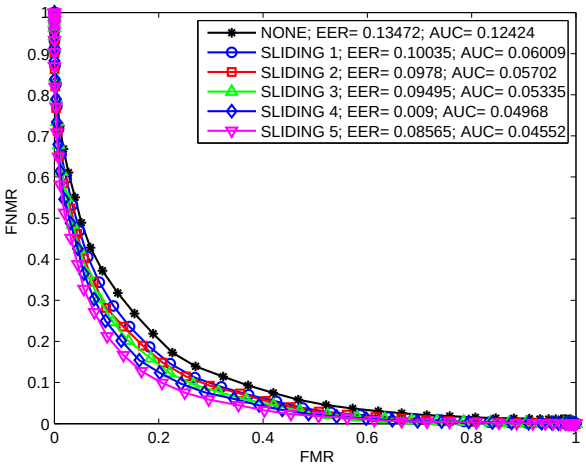
Varying the update thresholds from one session to another allows reducing the update error rates, so the performance gets better over time in comparison to using a fixed or individual threshold. The method has been validated on a template update system for keystroke dynamics on two datasets. We have shown that our approach (based on sliding or growing windows) gives a better performance than the classical ones (EER 2% lower).



(a) Global thresholds.



(b) Individual thresholds.



(c) Variable threasholds.

Figure IV.6 – Performances of sliding window update method applied on GREYC 2009 database.

IV.3.4.2 Template adaptation

The main contributions of the proposed template adaptation method are:
1) It is initialized with a single sample as a reference
2) A multi-distance classifier is considered with adaptive weights.
In fact, we propose a contribution in each of the five components of the template update approach, as depicted in Algorithm 8.

- **Reference modeling:** By initiating the authentication process, users are supposed to type their passwords only once for the computation of the reference template. Afterwards, they can test their identity verification. The main idea is to limit the enrollment phase to a strict minimum and to allow an adaptation of the biometric

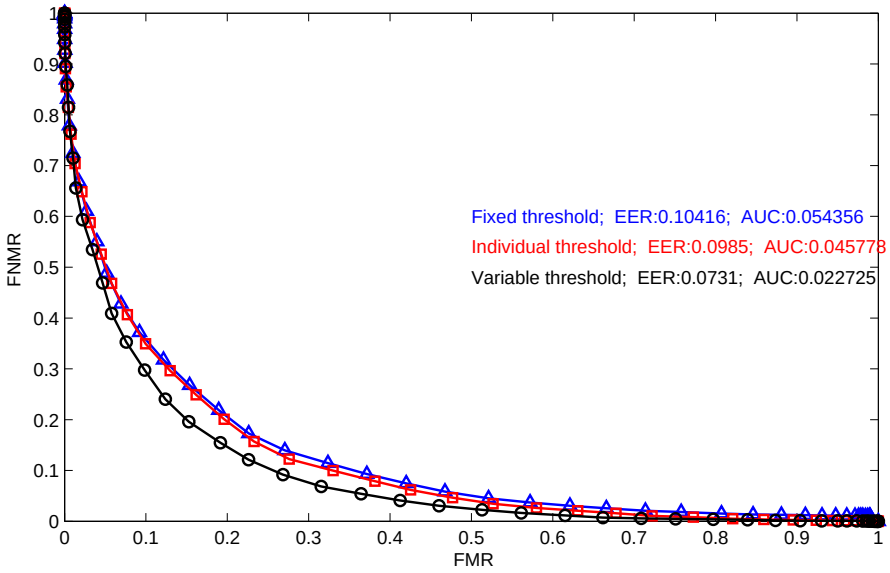


Figure IV.7 – Performances of growing window update with different thresholds tested on GREYC 2009 database

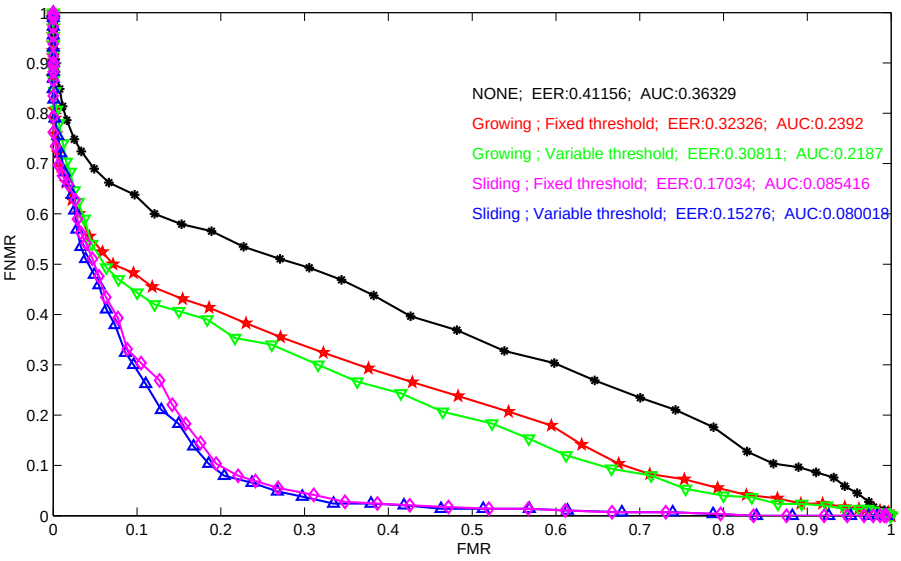


Figure IV.8 – Performances of growing and sliding window updates with different thresholds tested on Web-GREYC database

reference to fit its aging over time. Indeed, it is always mentioned that the enrollment phase annoys users [Giot et al., 2011b]. Even if the proposed scheme does not capture any variability during the enrollment stage, the combination with the proposed adaptation strategy will allow users to cope with it.

- **Adaptation criterion:** Different adaptation criteria have been used in the literature to initiate the adaptation process. Based on the double threshold mechanism, we put forward our new adaptation criterion called *"adapted thresholds"*. As demonstrated in section IV.3.4.1, it uses individual thresholds that are decreased over time according to Equation (V.3). In fact, after using the password for a long period, the intra-class variation in the user's keystroke dynamics becomes lower. This is due to the acquisition of a habit after frequent uses. Therefore, we slightly reduce the threshold during the use of the system.

- **Adaptation mode:** Adaptation is dealt with in a semi-supervised mode through the application of the "*GA-KNN verification method*". Each query is labeled with the KNN classifier. It will be accepted (*i.e.* classified as genuine) if the value of the global score $Score_j$, calculated according to Equation (IV.4), is lower than the *"adapted threshold"*. Equation (IV.4) permits calculating the weighted sum of the four partial scores (D_{STAT}, $D_{HAMMING}$, $D_{EUCLIDIAN}$, $D_{MANHATTAN}$) which are the nearest neighbor scores obtained by the KNN classification with four different distance metrics, defined by Equation (IV.3). The weight parameters (α, β, γ, δ) are calculated thanks to GA. Algorithm 8 details the process.

  GA is inspired by the natural evolution process following the Darwinian model. It uses a fitness function to optimize the weight parameters (α, β, γ, δ) during a number of iterations (or generations). The content of the initial population is randomly generated. For our experimentations, the optimization function of GA minimizes the FNMR and the FMR by optimizing the Half-Total Error Rate (HTER). The computation of the FNMR and FMR values is based on the presented queries for each adaptation session which are labeled thanks to the GA-KNN. The adopted parameters of the GA algorithm are summarized in Table IV.5.

Table IV.5 – GA Parameters

Parameter	Value
Population size	50 (number of variables $\leqslant$ 5)
Crossover fraction	0.8
Generation	400 (100*number of variables)
Elite count	2.5 (0.05 * population size)
Selection function	Stochastic uniform
Crossover function	Crossover scattered
Mutation function	Gaussian

We periodically vary the classification parameters (α, β, γ, δ) of Equation (IV.4) to ensure a better performance. Consequently, at the end of each adaptation session, the GA recalculates new global weights for all the users to optimize them. In each session, the fitness of all users is evaluated and is usually the value of the defined optimization function.

The great advantage of GA is that it succeeds in finding optimal solutions for very complex problems, so as to take advantage of certain known properties. Furthermore, they are used in applications where a large number of parameters are involved and where it is necessary to obtain good solutions in only few iterations in real-time systems, like in the suggested approach.

- **Adaptation periodicity:** The proposed adaptation strategy operates online. Each accepted query that satisfies the adaptation criterion is included in the adaptation mechanism.

- **Adaptation mechanism:** The initial reference is composed of only a single keystroke dynamics sample. Therefore, the suggested process enriches the reference describing the user's typing manner as shown in Figure IV.9. At the beginning, the growing window mechanism is adopted. As a result, each request accepted by the adaptation criterion is added to the gallery samples. Once the maximal size of the users' gallery (set to 10 samples in our work) is reached, the sliding window mechanism will be applied. Thereby, the oldest sample in the reference gallery will be replaced by a new query. Hence, the process is a "*double serial mechanism*".

We noticed that the "double serial mechanism" allows us to obtain a satisfying model of the users' typing rhythm evolution over time. In fact, the novelty is to combine the two considered adaptation mechanisms by applying them sequentially to the same reference. At first, the growing window mechanism is useful for increasing the number of samples representing the users' reference. The purpose of this phase is to enrich the description of the users' typing manner. After that, we update the reference to take into account the intra-class variations over time. Indeed, the sliding window mechanism starts when the size of the reference reaches 10 samples in order to keep a minimal size of the reference (no waste in calculation time). Moreover, this adaptation mechanism is based on the principle that the oldest samples are the least representative of the actual keystroke dynamics of the user. As demonstrated in Algorithm 7, the newest samples are added while the oldest ones are deleted. In the next section, we demonstrate the efficiency of the proposed method.

Algorithm 7: Proposed template update strategy for user j during an adaptation session.

Require:
$ref_{j_{(t)}}, q_j^s, maxSize(ref_{j_{(t)}}), N \leftarrow maxSize(ref_{j_{(t)}}), (\alpha_0, \beta_0, \gamma_0, \delta_0) = EmpiricalValues$
Ensure: $ref_{j_{(t+1)}}$

for `j = 1 : NumberOfUsers` **do**
  for `e = 1 : 8` **do**
    $D_{STAT} \leftarrow KNN_{Statistical}(ref_{j_{(t)}}, q_j^e, K = 1)$
    $D_{HAMMING} \leftarrow KNN_{Hamming}(ref_{j_{(t)}}, q_j^e, K = 1)$
    $D_{EUCLIDIAN} \leftarrow KNN_{Euclidean}(ref_{j_{(t)}}, q_j^e, K = 1)$
    $D_{MANHATTAN} \leftarrow KNN_{Manhattan}(ref_{j_{(t)}}, q_j^e, K = 1)$
    $Score_j = \alpha_s \times D_{STAT} + \beta_s \times D_{HAMMING} + \gamma_s \times D_{EUCLIDIAN} + \delta_s \times D_{MANHATTAN}$
    if ($Score_j < adaptatedThreshold$) **then**
      if ($size(ref_{j_{(t)}}) < N$) **then**
        $ref_{j_{(t+1)}} \leftarrow GrowingWindow(ref_{j_{(t)}}, q_j^e)$
      else
        $ref_{j_{(t+1)}} \leftarrow SlidingWindow(ref_{j_{(t)}}, q_j^e)$
      end if
    end if
  end
end

$\alpha_{s+1}, \beta_{s+1}, \gamma_{s+1}, \delta_{s+1} \leftarrow GeneticAlgorithm((\alpha_s, \beta_s, \gamma_s, \delta_s); (D_{STAT}, D_{HAMMING}, D_{EUCLIDIAN}, D_{MANHATTAN}))$

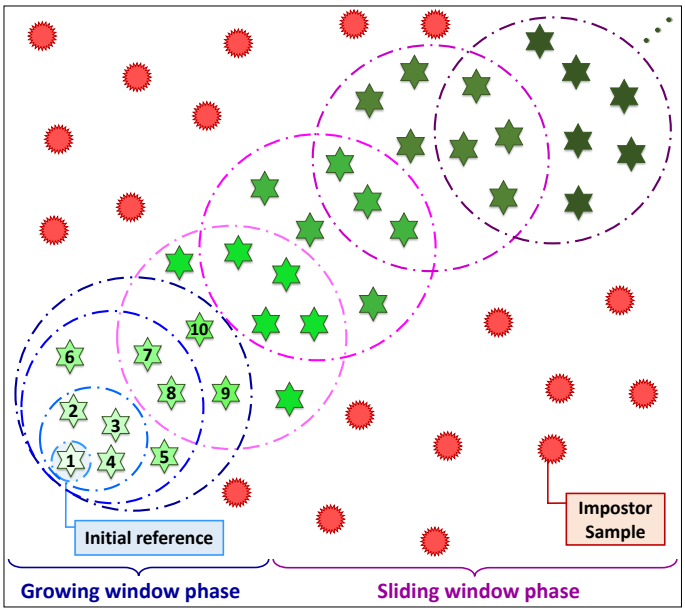


Figure IV.9 – User's gallery representation over time: The effects of the *double serial mechanism* on the gallery. Each circle represents the gallery samples in a specific session.

IV.4 Experiments

In this section, we put forward the processing description. Moreover, we present the evolution of some parameters of the experiments like the reference gallery size and the weight parameters over the adaptation sessions. We also detail the obtained results for each adaptation session.

IV.4.1 Data stream generation

To evaluate the performance of the proposed system and to follow its evolution, we divided the adaptation process into different sessions. For each session, we introduced eight new queries to the system. These data were divided into five genuine samples and three impostor ones for each adaptation session. First, 5 genuine queries are presented to the authentication process.They were presented according to the chronological order of the database safeguard. Subsequently, the three imposter queries were randomly introduced.

The biometric data stream was then divided into 37.5% (3/8) of impostor samples and 62.5% (5/8) of genuine samples. The attack rate was higher than that generally used in keystroke dynamics studies [Giot et al., 2012c, Pisani et al., 2016] (70% for genuine samples and 30% for impostor ones).

For both GREYC-2009 and GREYC-WEB databases, we have 60 samples for each user. These samples were divided into 12 sessions (5 $genuine\ samples/session$). As we used the first sample as initial reference, we presented in the last session only four genuine samples. The impostor attacks were randomly generated by the samples of other users of the database. For the CMU database, 400 samples per user are available. The system operates for 80 sessions.

IV.4.2 Classification parameters

In this work, we opted for a KNN classifier based on multi-distances. Thus, to set the values of the vote parameters (α_s, β_s, γ_s, δ_s) of Equation (IV.4), we used GA. It is a widespread algorithm that provides high-quality solutions for optimization problems. Its advantage is that it can start from a collection of randomly generated data. This is quite similar to our experimentation conditions, where the initial reference is not random, but it does not represent well the users (using only the first sample).

The initial values of the vote parameters are empirically set. We opted for the values that guarantee the best performances for the first adaptation session. At the end of each session,

Table IV.6 – Classification parameters obtained with GA during 12 sessions for GREYC 2009 database

Adaptation Sessions	Parameters			
	α	β	γ	δ
1	0.0381	0.6295	0.3327	-0.0002
2	0.0289	0.5662	0.3143	0.0906
3	0.0183	0.6298	0.2562	0.0958
4	0.0560	0.6437	0.2854	0.0149
5	0.0404	0.6534	0.2867	0.0196
6	0.0482	0.6024	0.3172	0.0322
7	0.0506	0.6924	0.1904	0.0667
8	0.0327	0.6581	0.2582	0.0511
9	0.0616	0.6684	0.2475	0.0225
10	-0.0593	0.6681	0.3743	0.0170
11	0.0374	0.6936	0.2732 -	0.0042
12	0.0468	0.6411	0.2460	0.0661

Table IV.7 – Classification parameters obtained with GA during 12 sessions for GREYC-WEB database

Adaptation sessions	Parameters			
	α	β	γ	δ
1	0.1611	0.4291	0.4201	-0.0103
2	0.1127	0.4612	0.3666	0.0595
3	0.1424	0.4427	0.3603	0.0546
4	0.1694	0.4669	0.3963	-0.0326
5	0.1325	0.4219	0.3754	0.0702
6	0.1425	0.4333	0.3368	0.0874
7	0.1369	0.3822	0.4118	0.0691
8	0.1191	0.4675	0.3841	0.0293
9	0.1453	0.3993	0.3875	0.0679
10	0.1078	0.4995	0.4102	-0.0175
11	0.1114	0.4926	0.03203	0.0757
12	0.1239	0.4366	0.3851	0.0544

after the presentation of 8 new queries, we restart the GA to update the weight parameters. These new parameters would guarantee minimal FNMR and FMR rates.

This process was repeated for each adaptation session. Table IV.6 and Table IV.7 present the weight values obtained at the end of each adaptation session for GREYC-2009 and GREYC-WEB databases, respectively. We can notice that the Hamming and Euclidean distances (β_s and γ_s) have in general the highest vote values. In fact, these two distances demonstrated also better performances than the others while testing each of them separately with the KNN classifier [Mhenni et al., 2018b].

IV.4.3 Gallery size

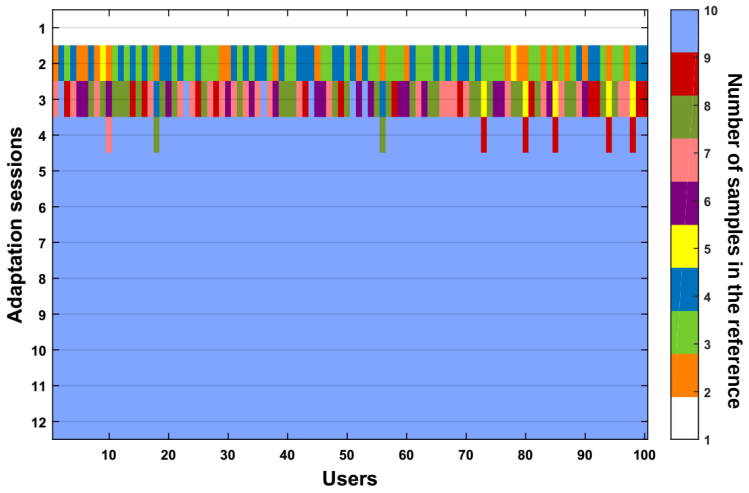
As previously mentioned, the initial reference gallery contains only one sample of the genuine user introduced during the enrollment phase. Throughout the adaptation strategy, the reference size increases over time by adding each accepted query to the user's reference. Once the maximal size (ten samples) is reached, the reference size will remain stable.

Since the number of accepted queries is not the same for all users, the size of the reference differs from one user to another at the end of the session. We followed this gallery size variation during all adaptation sessions to separate between sessions belonging to the growing window phase and those belonging to the sliding window one.

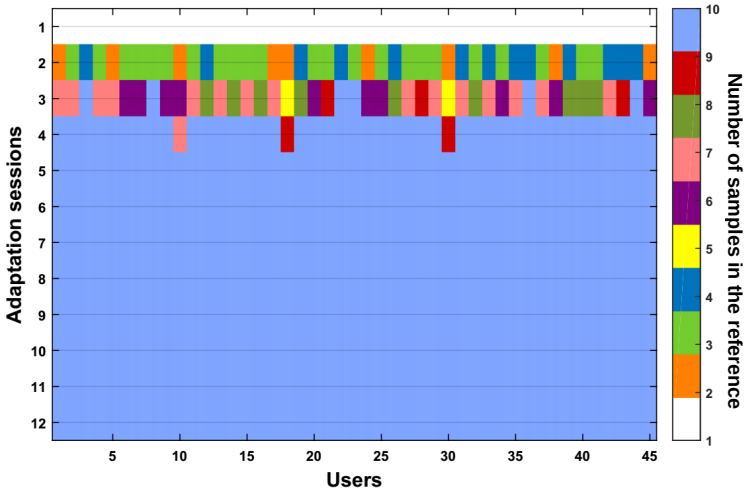
For a better illustration of the distribution of the database users according to their reference size, we added the representation of the reference size variation in a table forms. As depicted in Table IV.8, IV.9 and IV.10, the growing window phase operates over a limited number of sessions. Its duration does not exceed three sessions for some users and 5 sessions for all users. This implies that the number of false rejections is not high since the beginning. Our experimentations demonstrate that despite the lack of samples in the initial reference, the classifier can distinguish between novel queries. Few users have a reference size equal to 5 at the end of the first session (only 2 users for the GREYC-2009 database). For the following sessions, the number of genuine accepted queries goes up quickly especially for the CMU database as illustrated in Figure IV.10.

Table IV.8 – Size of references in the beginning of each adaptation session for GREYC 2009

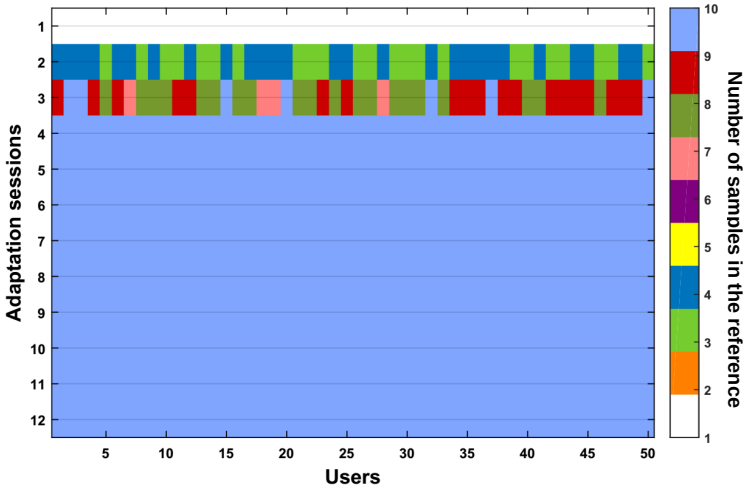
Reference size	1	2	3	4	5	6	7	8	9	10
Session 1	100%	-	-	-	-	-	-	-	-	-
Session 2	-	22%	43%	33%	2%	-	-	-	-	-
Session 3	-	-	-	2%	5%	18%	25%	30%	14%	6%
Session 4	-	-	-	-	-	-	1%	2%	5%	92%
Session 5-10	-	-	-	-	-	-	-	-	-	100%



(a) GREYC 2009 database



(b) GREYC-WEB database



(c) CMU database

Figure IV.10 – Size variation of the users' references size during all adaptation sessions

Table IV.9 – Size of references in the beginning of each adaptation session for GREYC-WEB

Reference size	1	2	3	4	5	6	7	8	9	10
Session 1	100%	-	-	-	-	-	-	-	-	-
Session 2	-	20%	51.1%	28.9%	-	-	-	-	-	-
Session 3	-	-	-	-	4.5%	22.2%	33.3%	20%	6.7%	13.3%
Session 4	-	-	-	-	-	2.2%	-	-	4.4%	93.4%
Session 5-10	-	-	-	-	-	-	-	-	-	100%

Table IV.10 – Size of references in the beginning of each adaptation session for CMU

Reference size	1	2	3	4	5	6	7	8	9	10
Session 1	100%	-	-	-	-	-	-	-	-	-
Session 2	-	-	46%	54%	-	-	-	-	-	-
Session 3	-	-	-	-	-	-	8%	40%	38%	14%
Session 4-80	-	-	-	-	-	-	-	-	-	100%

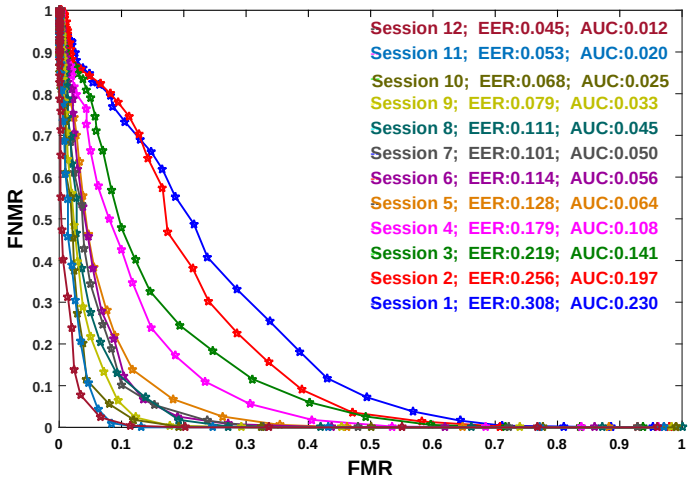
IV.5 Experimental results and discussion

In the experimental results, the following performance metrics were adopted: FNMR, FMR, EER, AUC and the Accuracy. We chose these performance measures in order to compare the obtained results with other studies in the literature.

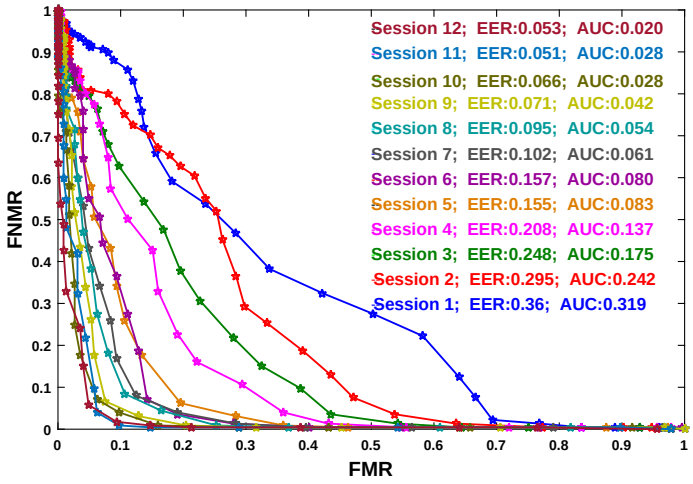
Figure IV.11 depicts the EER and AUC values of each adaptation session for the three considered databases. Concerning the CMU database, as the number of sessions is quite high (80 sessions), we illustrate only the performances of every ten sessions. We can conclude that the results are slightly improved in each session. The performance improvements during the sliding window phase are much clearer than those of the growing window one. These performances are expected since the reference is not entirely defined at the beginning.

The final result of the last session illustrates a statistically significant improvement. The obtained performances (EER, AUC) in the last session are much improved compared to those of the first one, as shown in Figure IV.11.

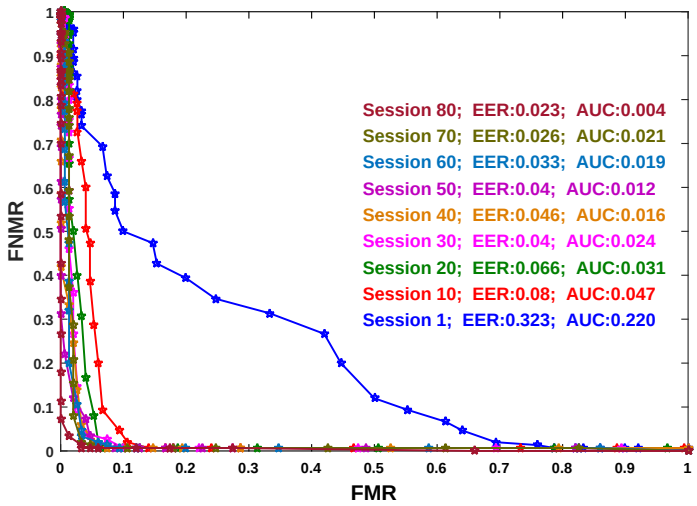
As the proposed method is processing online, we are interested in the computation time of each phase. Table IV.11 presents the computation time of each phase for a single user and for all considered users of the GREYC 2009 database. Concerning the computation time of a unique user, the average computation time is considered. Timing is calculated on CPU with



(a) GREYC 2009 database



(b) GREYC-WEB database



(c) CMU database

Figure IV.11 – DET curves and associated performance results (EER, AUC) for all adaptation sessions.

an Intel i7 processor with a speed of 2.5 GHz and 8-Gb RAM. The adaptation phase is faster than the other steps of the process. All phases have a fast computing time except GA which operates in an offline way, so it does not affect the operating time of the proposed approach.

Table IV.11 – Computation time in seconds involved in each phase of process for only one user and for all users

Phase	One user	All users
Feature extraction	0.035	4.53
Pre-processing: aberration	0.012	0.98
Pre-processing: normalization	0.000015	0.0016
Enrollment	0.00009	0.0015
Verification: Statistical	0,006	0,63
Verification: Hamming	0.002	0.187
Verification: Euclidean	0.0015	0.158
Verification: Manhattan	0,0017	0,176
Genetic algorithm	-	18.93
Adaptation	3.2395e-05	0.0012

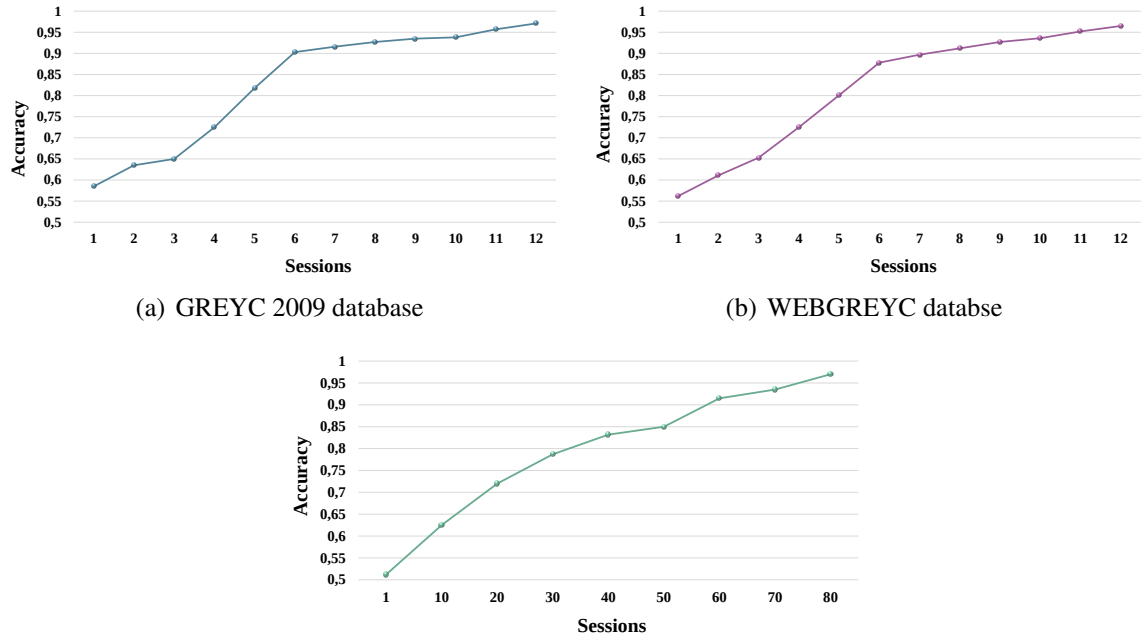
The overall results of FMR, FNMR and accuracy concerning the three considered databases are shown in Table IV.12. These results are calculated over all adaptation sessions while considering the whole data of the databases. The best achieved results are those obtained with the GREYC-2009 database. While considering the EER and AUC performances, the CMU database presents the best obtained results.

Table IV.12 – Overall performances for three considered databases

	GREYC-2009	GREYC-WEB	CMU
FMR	0.0833	0.1375	0.1406
FNMR	0.0463	0.0516	0.0647
Accuracy	0.828	0.810	0.794

Actually, we have considered two metrics to evaluate the proposed approach, EER and AUC, since they have been commonly used. Compared to previous works using the same evaluation metrics and applied to the same database, the proposed method performs better as illustrated in Table IV.13. We have also compared our approach to the enhanced template update [Pisani et al., 2016] applied to the same database, but the considered evaluation metric was the accuracy. For that purpose, we calculated the accuracy corresponding to all

adaptation sessions. Even if it is a bit low, the obtained accuracy is better than that of the enhanced template update.



(a) GREYC 2009 database

(b) WEBGREYC databse

(c) CMU database

Figure IV.12 – Accuracy during all adaptation sessions for the three considered databases.

In the beginning of the process, the obtained accuracy was a bit low. However, it improved over time, and it was quite high in the final sessions. We represented, in Figure IV.12 the accuracy variation during all adaptation sessions for the 3 considered databases.

We compared the proposed method to previous works in the literature. For GREYC-2009 database, the proposed method was compared to the average mechanism [Giot et al., 2011b], which was applied to a reference initially composed of a gallery with five samples and not exceeding 15. For that, we investigated various threshold types: global, individual and variable. Once again, the adaptation mechanism based on variable thresholds led to better performances than when using the other types of thresholds. Table IV.13 shows the comparison of the obtained results on the GREYC-2009 database.

We also compared the proposed method with some other work from the literature to analyze the impact of the number of samples used in the reference gallery, especially in the training phase (See Table IV.14). For the CMU database, the best obtained result by the Enhanced template update in [Pisani et al., 2016] was 0.670 accuracy, although the reference was obtained by 40 samples. For our experiments, the results achieved with the suggested method were much better with a unique sample as an initial reference template as we obtained 0.794 accuracy.

Table IV.13 – Comparison of obtained results with different thresholds for GREYC-2009 database

Threshold	Double serial mechanism		Average mechanism	
	Reference size	EER%	Reference size	EER %
Global	1-10	7.1	5-15	6.96
Individual	1-10	6.5	5-15	6.95
Variable	1-10	4.5	-	-

Table IV.14 – Performance comparison of the different implemented mechanisms

Database	Adaptation Mechanism	Gallery size	Threshold	FNMR	FMR	Accuracy
CMU	Double serial mechanism	1-10	Variable	0.064	0.140	0.794
CMU	Enhanced template update	40	Global	0.088	0.573	0.670
WEB-GREYC	Double serial mechanism	1-10	Variable	0.051	0.137	0.810
WEB-GREYC	Enhanced template update	40	Global	0.042	0.355	0.802

The proposed method had the advantage of minimizing the computation time to create the reference that was very important for the online adaptation mechanism. The experimental results showed that the obtained performance (EER, FNMR) outperformed the other methods in the state of the art for the same databases and under the same test protocol conditions. Furthermore, the proposed method satisfied the suggested conditions in an industrial context. Indeed, a single sample was necessary during the enrollment step. It was a great advantage instead of asking users to type their password multiple times.

To illustrate the advantages of the proposed adaptation approach, we applied other algorithms of the literature to the GREYC-WEB database. We firstly tested the growing window mechanism with a reference containing a single sample initially. The size of the reference increases up to 43 thanks to the adaptation mechanism. Secondly, we applied the sliding window mechanism based on a 10-sized reference. Thirdly, we also tested the proposed double serial mechanism while the reference was initialized to 5 samples and its maximum size was fixed to 10. Finally, the double parallel mechanism was conducted using two sub-references. One of them initially contained a single sample and it was adapted with the growing window mechanism. The other one initially comprised 10 samples and it was adapted with the sliding window mechanism. Figure IV.13 depicts the size variations for each adaptation mechanism. All of these mechanisms were implemented by the GA-KNN method based on the weighted vote of 4 distance metrics. The obtained results are summarized in Figure IV.14.

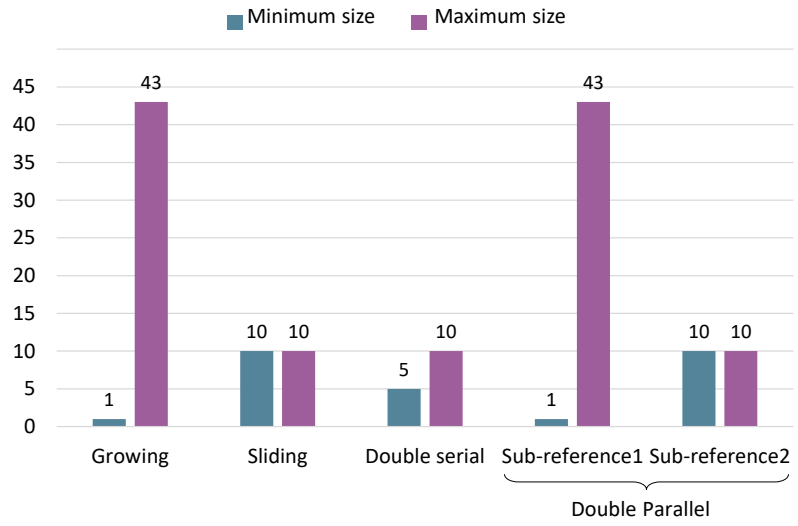


Figure IV.13 – Minimum and maximum reference size for compared mechanisms.

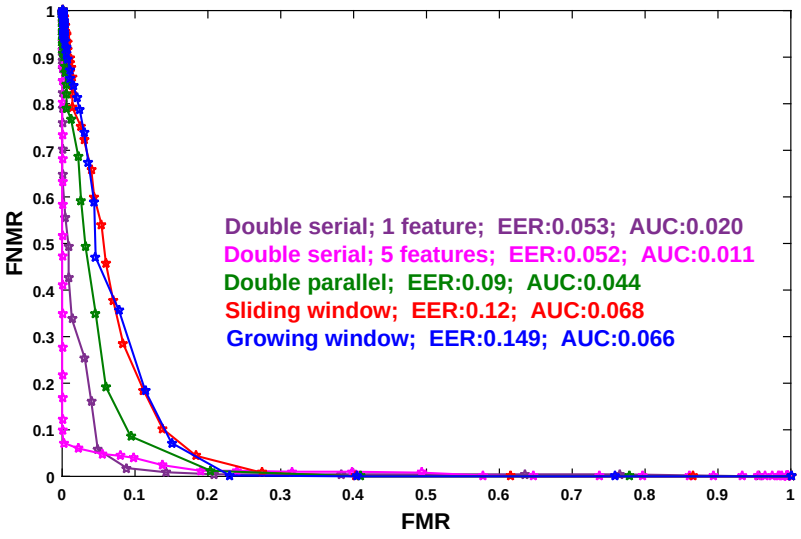
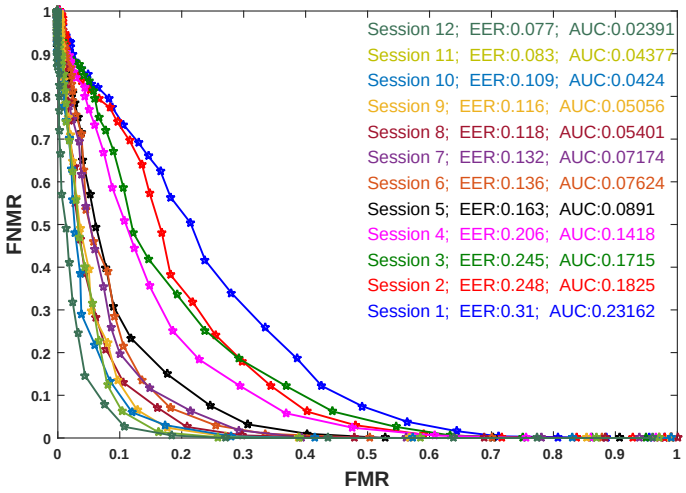
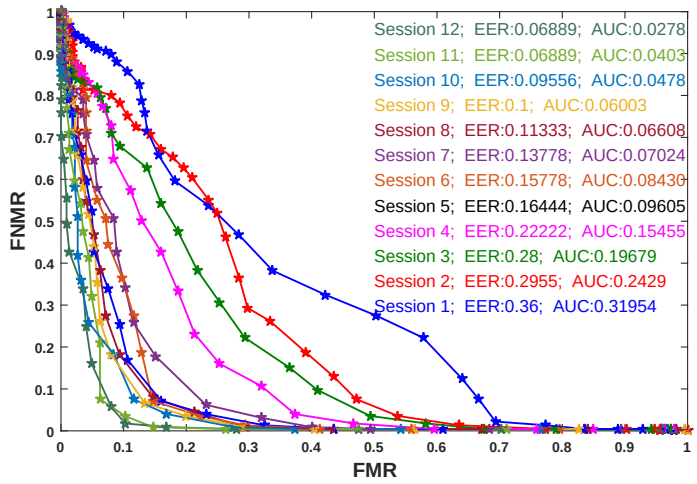


Figure IV.14 – DET curves of last adaptation sessions and associated performances (EER, AUC) of different adaptation mechanisms applied to GREYC-WEB database

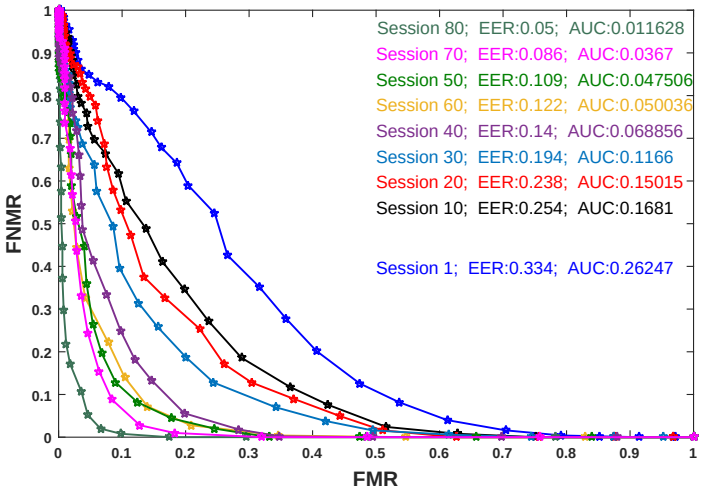
With a reference size approximately equal to the proposed approach, the double serial mechanism was the best performing among the tested mechanisms. While increasing the initial size of the reference by five samples, we obtained better performances. This was due to the larger description of the keystroke dynamics of users. However, the performance difference at the final session was not very large. Thus, we chose an approach based on a single sample in the learning phase in order to familiarize it with the industrial application environment.



(a) GREYC 2009 database



(b) WEBGREYC databse



(c) CMU database

Figure IV.15 – DET curves and associated performance results (EER, AUC) for all adaptation sessions while considering 2 majority votes.

The best performances achieved was obtained by consiedring the vote of the 4 chosen distance metrics. Indeed, we tested the proposed method on the three used databases considering only the two majority votes. In fact, Beta and Gamma are the most significant vote coefficients corresponding to Hamming and Euclidean distances respectively. The obtained performances were not improved as depicted in Figure IV.15. For CMU database for example, the best EER value obtained in the last adaptation session is equal to 0.05% while considering only two distances whereas it is equal to 0.02% while considering the 4 chosen distances.

IV.6 Conclusion

Adaptive biometric strategies provide an important solution to remedy the intrinsic intra-class variations in behavioral biometric authentication systems. As the keystroke dynamics is a biometric modality that suffers from continuous variations over time, adaptive methods are a good solution to compensate for this trouble. Most of the existing studies have used a huge number of samples to create the reference describing the users' typing rhythm in the enrollment phase.

This chapter has investigated a solution that enables modeling an individual's keystroke dynamics while minimizing the used samples for the definition of the reference template. For this purpose, we proposed a single enrollment process (the password was typed only once during the enrollment step). The size of each user gallery would increase while using the system, to reach a maximum size equal to 10 thanks to the *double serial mechanism*. Actually, the growing window first serves to enlarge the users' galleries so as to capture more intra-class variability. When the maximum size of the reference is attained, the sliding window will take place and allow following the temporal variation in the users' keystroke dynamics. The proposed contribution is an interesting solution as it satisfies industrial needs (usable enrollment and good efficiency).

We also evolved a *GA-KNN verification method* to achieve better performances during the whole adaptation session. Indeed, the weights from different distances for the KNN classifier, in addition to the GA optimization, are useful to minimize recognition errors. With regards to previous work, the suggested method shows a great performance improvement. As it has been applied on several databases, it has demonstrated competitive performances in each database.

To improve the performances of the proposed method, we decided to put forward an adaptive strategy to each category of users according to their specificities. In fact, we noticed that user's behaviors over time are quite different and can be divided into groups. Thus, we

considered "Doddington zoo" to classify them and to apply an adaptive strategy appropriate to each users' class.

CHAPTER V

Adaptive Biometric Strategy using Doddington Zoo Classification

V.1 Introduction

Regarding the fact that individuals have different interactions with biometric authentication systems, several techniques have been developed in the literature to model different users categories. Doddington Zoo is a concept of categorizing users behaviors into animal groups to reflect their characteristics with respect to biometric systems. This concept was developed for different biometric modalities including keystroke dynamics. The present study extends this biometric classification, by proposing a novel adaptive strategy based on the Doddinghton Zoo, for the recognition of the user's keystroke dynamics. The obtained results demonstrate competitive performances on significant keystroke dynamics datasets.

V.2 Doddington zoo theory

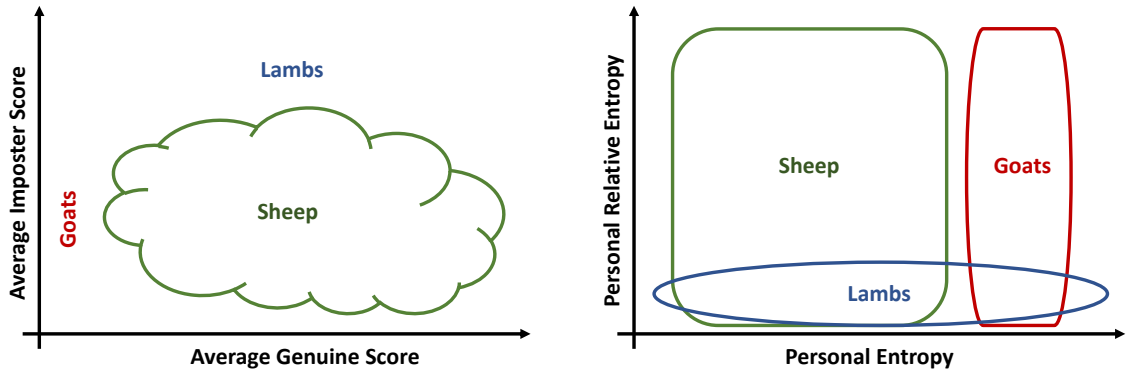
Keystroke dynamics is a behavioral modality non intrusive, inexpensive and weakly constrained for the user [Rybnicek et al., 2014, Mhenni et al., 2016].

The major drawback of this modality is that it suffers from large intra-class variation [Epp et al., 2011, Nahin et al., 2014]. In fact, the keystroke dynamics of the user varies as time elapses according to different situations. This variability may be due to the familiarity with the password after a time span, the user's humor and activeness and the changing of the keyboard (AZERTY or QWERTY, virtual or physical).

Adaptive strategies [Didaci et al., 2014, Poh et al., 2012] also known as template update strategies are an interesting solution to overcome the intra-class variability. Commonly, a unique adaptation mechanism is applied to all users of the authentication process. Meanwhile, it was demonstrated that biometric systems performances are subject dependent [Poh et al., 2015a]. That is why, we decided to use an update strategy for each category of users in this work. For that purpose, we are interested in the users classification based on the Doddington Zoo [Doddington et al., 1998]. It is a widely used theory for user classification [Ross et al., 2009, Morales et al., 2014], but, to our knowledge, it has not been mixed with adaptive strategies for keystroke dynamics modality. The common users' classes are :

- sheep: users who can easily be recognized;
- goats: users who are particularly difficult to recognize;
- lambs: users who are easy to imitate;
- wolves: users who can easily imitate others.

Several methodologies have been proposed to distinguish between this variety of users as shown in Figure V.1. Doddington *et al.* considered the classification based on the mean of the user's genuine or impostor scores. Indeed, users classified as goats increase the False Non Match Rate (FNMR) of the recognition system whereas wolves and lambs increment its False Match Rate (FMR). Other research works [Houmani & Garcia-Salicetti, 2016] proposed to use the personal entropy and relative entropy for biometric menagerie of online signature verification. Personal entropy is computed using only genuine data. It serves to differentiate between sheeps and goats class of users. Relative entropy is calculated with both genuine and impostor data. It helps to distinguish lambs class.



(a) FNMR and FMR based classification. (b) Entropy based classification.

Figure V.1 – Animal groups of the Doddington ZOO biometric menagerie according to [Houmani & Garcia-Salicetti, 2016].

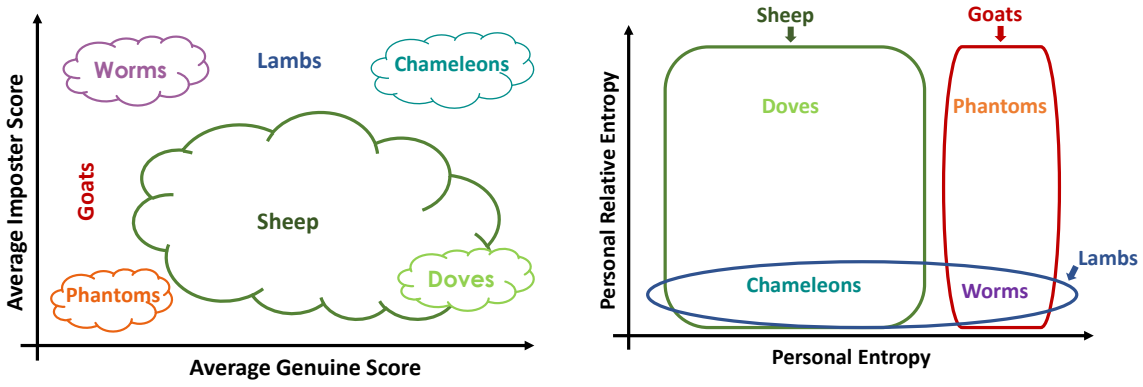
Besides, sheeps generally dominate the population of the zoo, goats as well as lambs constitute only a small fraction of the population. However, the wolves category constitutes a large portion of false rejection and acceptance rates.

Further, Yager and Dunstone [Yager & Dunstone, 2007] distinguished four other animal categories of users by considering simultaneously both the genuine and impostor matching scores, for each claimed identity:

- Chameleons: corresponds to users who are easy to recognize and easy to attack;
- Phantoms: depicts the users characterized by rejections of genuine and impostor queries;
- Doves: represents the best users because they are easy to recognize and difficult to attack;
- Worms: regroup the worst users as they are difficult to recognize and easy to attack.

The four additional sub-categories can also be distinguished thanks to the FMR and FNMR based classification or the entropy based classification as depicted in Figure V.2. For the FMR and FNMR based classification, chameleons belong to the users that are known by high genuine and impostor match scores. Contrariwise, phantoms are characterized by low genuine and impostor match scores. Doves are a sub-group of Sheep according to this classification methodology. They are the best users since they lead both to high genuine and low impostor match scores. Worms in the opposite, are a sub-group of Goats. They represent the worst users, as they lead to low genuine and high impostor scores. This categorization was applied to diiferent modalities like face, speech, fingerprint, iris and keystroke modalities [Yager & Dunstone, 2010], but it was not associated to an adaptive strategy specific to each category of user.

The second method is to distinguish between these classes by the entropy based classification. First, chameleons are a sub-category of goats and lambs as they are known by the lowest Personal Entropy and the lowest Relative Entropy. Second, phantoms are a sub-category of goats class regarding that they have a reference with poor data quality generated in the enrollment phase. They are characterized by a low Personal Entropy and a low Relative Entropy. Third, doves are a sub-category of sheep class. They are characterized by the lowest Personal Entropy and the highest Relative Entropy. Finally, worms are a sub-category of goats and lambs classes. They have the highest Personal Entropy and the lowest Relative Entropy.



(a) FNMR and FMR based classification.

(b) Entropy based classification.

Figure V.2 – Large animal groups of the Doddington ZOO biometric menagerie according to [Houmani & Garcia-Salicetti, 2016].

In this chapter we are interested in the entropy based clssification to distinguish between the users characteristics. For that purpose, we examined the entropy of the users of the WEBGREYC database [Giot et al., 2012a] over time. We calculated the entropy of each

user's set of 5 samples in chronological order of the database safeguard. As depicted in FigureV.3, the characteristics of some users are stable over time such as those of user 3 and user 30. Others have an entropy that decreases over time like user 34. This means that their intra class variation decreases thanks to the mastery of the password for example. However, user 4 and user 11, have an increasing entropy. Their intra class variations increase due to different parameters like their emotional state. Thus the need for a user specific adaptation strategy is clearly demonstrated.

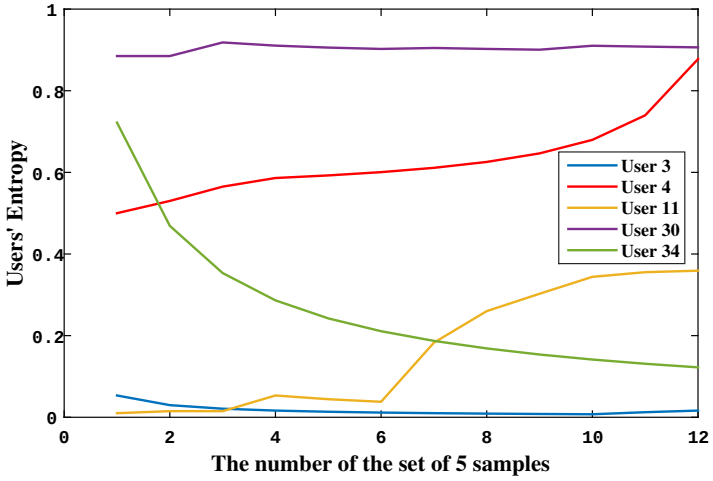


Figure V.3 – Personal entropy of some users of WEBGREYC database.

The proposed contributions investigate an authentication method based on a single capture of the user's keystroke dynamics in the enrollment phase in coherence with the single enrollment adaptation strategy proposed in the previous chapter IV. During the use of the authentication system, the reference is enriched thanks to the user dependent adaptive strategy. Users classification into Doddington Zoo categories is firstly based on the evolution of the user's reference size over time. Once the fixed maximum size of the reference is reached, the users categorization is ensured with the personal and the relative entropy calculation. Then, the adaptation strategy becomes specific to each category of users.

V.3 Three categories user specific adaptive system

This section presents a user dependent adaptive strategy according to the category of the Doddington zoo to which the user belongs.

V.3.1 Proposed adaptive strategy

The main idea consists in grouping users according to their performance evolution over time. Then, we put forward an adaptive strategy specific to each category of users, to ensure the usability of the keystroke dynamics modality. In the present work, three categories among the animal based categories are considered: sheeps, goats and lambs. Wolves class has been eliminated because we are not interested in modeling hackers in this work.

For the enrollment phase a single sample is considered to store the typing manner characteristics of each user. Afterwards, during the utilization of the authentication systems, the presented queries are classified based on the K Nearest Neighbor (KNN) classifier with multiple distances [Mhenni et al., 2018b]. These distances were chosen based on a comparison as explained in the previous chapter. Before taking the acceptance decision, a GA vote is performed [Mhenni et al., 2019b], based on all obtained scores. If the query is classified as genuine, it is used to update the user's reference. This process is achieved in an online way according to the algorithm 8.

Three adaptive mechanisms are considered for our process. The growing window mechanism is firstly used when the maximum size of the reference is not yet reached. Once the size of the user's reference is equal to the fixed maximum size, the sliding window mechanism is launched. Otherwise, the least frequently used mechanism is considered when the size of the user's reference is higher than the fixed maximum size. This is the case where the user migrates from the class of goats to that of sheep. The least frequently used mechanism is chosen to decrease the size of the reference from 15 to 10. Thus, the 5 least frequently used samples of the reference are deleted.

Some parameters and choices of the strategy need to be redefined and updated during the system's operation. So, we divided the process into sessions. Each session consists in the presentation of 8 new queries: 5 genuine queries and 3 impostor ones. The choice is similar to that of previous experiments shown in section IV.4.

At the end of each session, a parameters' adjustment is performed to optimize performance and ensure smooth operation, through 3 steps detailed in the following:

- *Users are assigned one of the three defined categories according to their characteristics:* During the growing window phase, the size of the reference is an important indicator regarding the category of the user. Indeed, if the size of the reference of the user remains small, this means that the number of accepted queries is limited. These users belong to the category of goats which are known as being difficult to recognize. The other part of the users, can be considered belonging to the sheep category since they are easily recognized.

Algorithm 8: Template update strategy for user j during an adaptation session.

Input : $ref_{j_{(t)}}$, $\mathcal{A}$ = {q}, $\theta_j^{adapt} = \left\{label^p, \text{maxSize}(ref_{j_{(t)}})\right\}$
Output : $ref_{j_{(t+1)}}$

nq←0 Number of accepted queries during the session
N ← size($ref_{j_{(t)}}$)
$score_1$ ← similarityScore $\left(KNN_{Hamming}(ref_{j_{(t)}});\mathbf{q}\right)$
$score_2$ ← similarityScore $\left(KNN_{Euclidean}(ref_{j_{(t)}});\mathbf{q}\right)$
$score_3$ ← similarityScore $\left(KNN_{Statistical}(ref_{j_{(t)}});\mathbf{q}\right)$
$score_4$ ← similarityScore $\left(KNN_{Manhattan}(ref_{j_{(t)}});\mathbf{q}\right)$
$score_j = \alpha \times score_3 + \beta \times score_1 + \gamma \times score_2 + \delta \times score_4$
if ($score_j$ <*adaptatedThreshold*) **then**
    nq←**nq**+1
    if (*N*<*maxSize*($ref_{j_{(t)}}$)) **then**
        $ref_{j_{(t+1)}}$ ← GrowingWindow ($ref_{j_{(t)}}$,**q**)
    else
        if (*N*==*maxSize*($ref_{j_{(t)}}$)) **then**
            $ref_{j_{(t+1)}}$ ← SlidingWindow ($ref_{j_{(t)}}$,**q**)
        else
            $ref_{j_{(t+1)}}$ ← LeastFrequentlyUsed($ref_{j_{(t)}}$,**q**)
        end
    end
end

However, during the other adaptation phases (sliding window and least frequently used mechanisms), the distinction of the users' categories is based on the Entropy measure, since the size of the reference is maximum and is no longer significant. First, the Personal Entropy is measured by means of local density estimation according to equation (V.1).

$$PersonalEntropy_j = -\sum_{i=1}^{N} ref_{j_{(t)}}(i)\ \log\left(ref_{j_{(t)}}(i)\right) \tag{V.1}$$

Where N is the number of samples in the reference.

If the Personal Entropy is low, then the user is classified as a sheep. Otherwise the user is considered as a goat. Additionally, the objective of this work requires assessing the vulnerability of a user to attacks. For this reason, we consider another quality

measure, namely Relative Entropy, which allows a user to be characterized not only in terms of keystroke dynamics variability, as Personal Entropy does, but also in terms of how difficult it is to attack such a typing manner. In fact, Relative Entropy defined in Equation (V.2), aims to recognize users belonging to lambs class.

$$EntropieRelative_j = \frac{1}{2}\left(\sum_{i=1}^{N} ref_{j_{(t)}}(i) \log\left(\frac{ref_{j_{(t)}}(i)}{attaq_j(i)}\right) + \sum_{i=1}^{N} attaq_j(i) \log\left(\frac{attaq_j(i)}{ref_{j_{(t)}}(i)}\right)\right) \tag{V.2}$$

where $attaq_j$ is a matrix containing N samples of the keystroke dynamics of multiple users other than the user j.

If the value of this entropy is low, the user is considered more vulnerable to attacks. Thereby, he/she is classified as a lamb. The details of this process are described by the algorithm 9.

Algorithm 9: Assign users to specific classes at the end of the session.

Input : $ref_{j_{(t)}}$, $attaq_j$
Output : goatsClass, sheepClass, lambsClass

if (N<$maxSize(ref_{j_{(t)}})$) **then**
  if ***nq***<3 **then**
    goatsClass← goatsClass ∪ $\{user_j\}$
  else
    sheepClass← sheepClass ∪ $\{user_j\}$
  end
else
  PE_j ← PersonalEntropy ($ref_{j_{(t)}}$)
  if PE_j<0.4 **then**
    sheepClass← sheepClass ∪ $\{user_j\}$
  else
    goatsClass← goatsClass ∪ $\{user_j\}$
  end
  RE_j ← RelativeEntropy ($ref_{j_{(t)}}$),$attaq_j$
  if RE_j<6 **then**
    lambsClass ← lambsClass ∪ $\{user_j\}$
  end
end

- *The vote parameters are controlled:* The parameters $(\alpha, \beta, \gamma, \delta)$ are generated by the Genetic Algorithm (GA) based on its parameters detailed in Table V.1. At the end of the first update session, after the creation of the three categories of users, these

parameters are generated to each category of users separately. At the end of each session, the vote parameters are updated thanks to the GA to fit the variation of each category population.

Table V.1 – Parameters of the Genetic Algorithm

Parameter	Value
Population size	50 (number of variables⩽ 5)
Crossover Fraction	0.8
Generation	400 (100*number of variables)
Elite count	2.5 (0.05 * population size)
Fitness Function	Minimizing the False Rejection Rate (FRR) and the False Acceptance Rate (FAR)
Selection Function	Stochastic uniform
Crossover Function	Crossover Scattered
Mutation Function	Gaussian

- *The used thresholds are updated:* The thresholds of acceptance and adaptation decision are adapted according to Equation (V.3). These thresholds are individual and adapted from one update session to another as defined in our work [Mhenni et al., 2016] and detailed in the previous chapter in section IV.3.4.1.

$$T_j^{i+1} = T_j^i - e^{-\frac{\mu_j}{\sigma_j}} \tag{V.3}$$

For all user categories, we changed some parameters according to the specificities of the user's category as summarized in Table V.2.

Table V.2 – Specific parameters according to user's category

User category	Reference size	Thresholds
Sheep	10	Adapted thresholds
Goats	15	Adapted thresholds
Lambs	10	Stricter thresholds

These choices are not arbitrary, the variation of the size of the reference and the chosen thresholds are justified as follows:

- For the sheeps class, standard settings are specified. The maximum size of the reference is fixed to 10, and the thresholds are adapted according to Equation (V.3).
- For the goats class, which is characterized by high intra-class variability, the description of their typing manner needs to be richer than that of other categories. For that, we have increased the maximum size of the reference of this user class to 15.
- Concerning the lambs, which are the most susceptible to be attacked as they are easy to imitate, stricter thresholds for the selection of new queries is the appropriate strategy. Thus, the thresholds of acceptance and update decision are updated according to Equation (V.4).

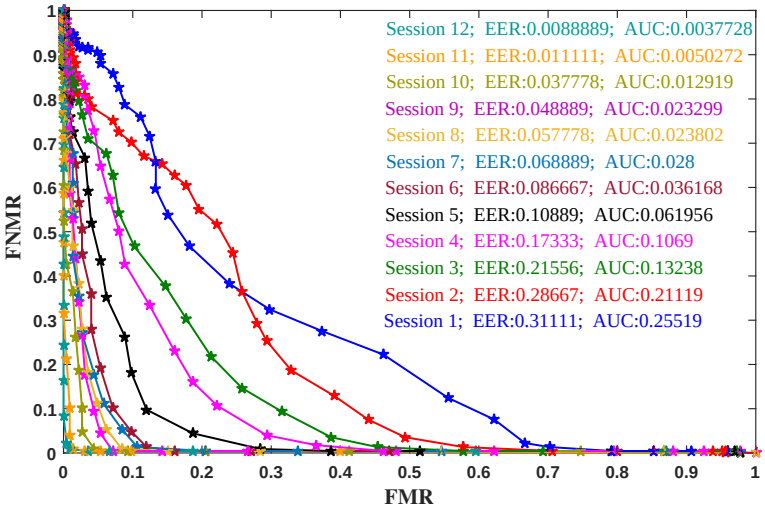
$$T_j^{i+1} = T_j^i - e^{-\frac{\mu_j}{2*\sigma_j}} \tag{V.4}$$

The biometric menagerie served to adjust all the parameters of the adaptive strategy (adaptive mechanism, reference size, decision thresholds, vote parameters, etc) to the users specificities, thus demonstrating competitive and promising performances as shown in the next section.

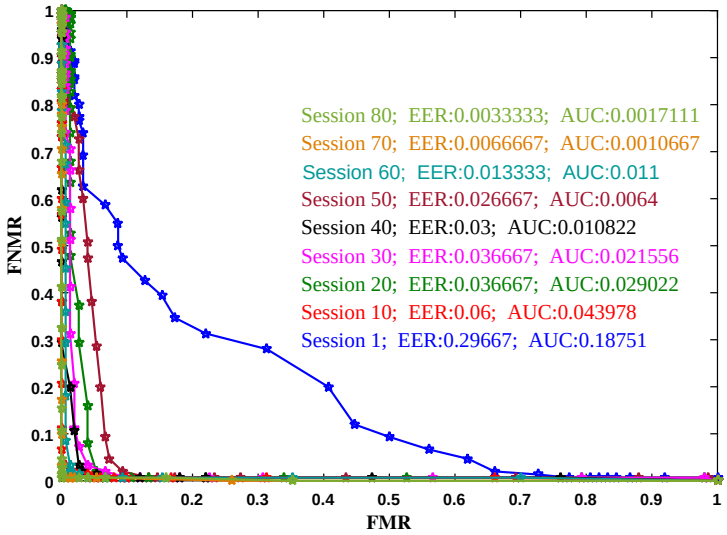
V.3.2 Experiments and results

The proposed approach is validated on two public datasets. In the WEBGREYC database [Giot et al., 2012a], 45 users participated in five acquisition sessions, typed the same password "SÉSAME" and provided 60 patterns. The CMU database [Killourhy & Maxion, 2010] includes the data of 30 users that typed the same password 400 times during eight acquisition sessions. The imposed password is ".tie5Roanl". Thus, for our experiments, we obtain 12 adaptation sessions for the WEBGREYC database (60/5) and 80 adaptation sessions for the CMU database (400/5).

To evaluate the performances of the proposed method, we used the Error Equal Rate (EER) and the Area Under Curve (AUC) metrics. The DET curves for the two considered databases are depicted in Figure V.4. The achieved performances are promising as the EER of the last adaptation session of the WEBGREYC and CMU database is respectively equal to 0.8% and 0.3%. Furthermore, we illustrate the variation of the size of users' references during the use of the system in Figure V.5. It is an indicator of the users' categories during the growing window mechanism as it represents how ease of recognizing the user.



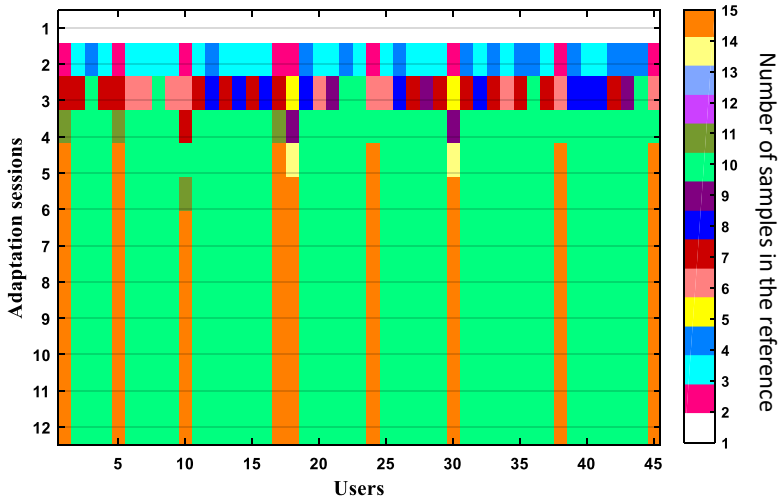
(a) WEBGREYC database.



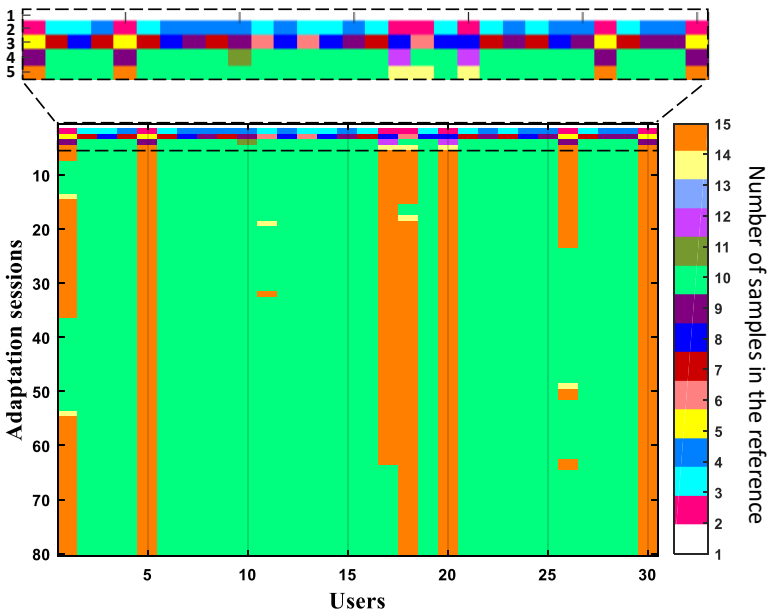
(b) CMU database.

Figure V.4 – Illustration of DET curves and the associated EER and AUC performances of each adaptation session.

The distribution of users categories among all adaptation sessions is also illustrated in Figure V.6. The sheep class represents the majority of users for both considered databases. Goats class represent approximately 0.2% for WEBGREYC and CMU databases. Lambs class represents 0.15% for WEBGREYC database and 0.16% for CMU database. Some users, especially for CMU database, switched from goats class to sheeps class thanks to the improvement in their performances.



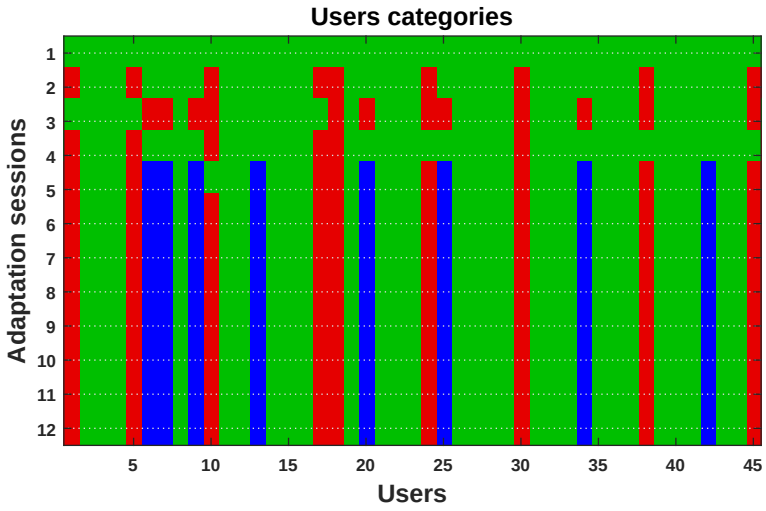
(a) WEBGREYC database.



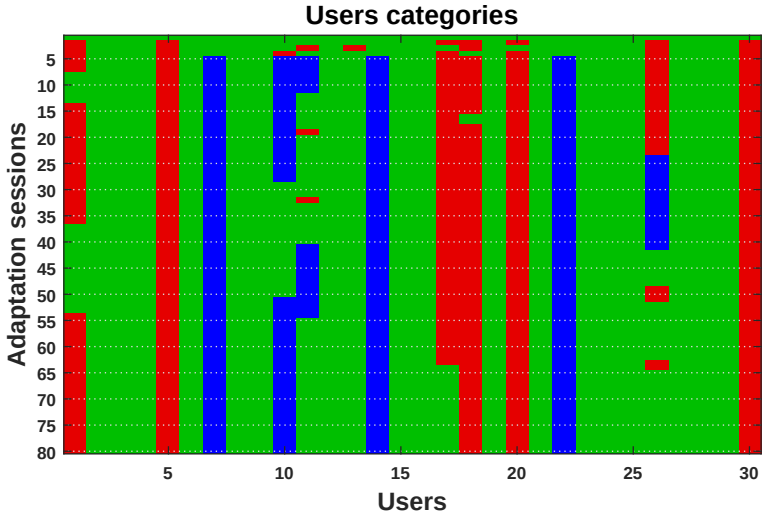
(b) CMU database.

Figure V.5 – Size variations of users' galleries during all adaptation session.

To highlight the advantages of the proposed method, we elaborate a comparison, depicted in Table V.3, of the proposed adaptation strategy with and without the biometric menagerie of Doddington Zoo. The achieved results demonstrate that classifying and updating users



(a) WEBGREYC database.



(b) CMU database.

Figure V.6 – Distribution of users categories during all adaptation sessions. The green color illustrates the sheep class, the red color illustrates the goats class and the blue color illustrates the lambs class.

reference according to their performances during the use of the system improved the EER performances by more than 2% going up to 4.5%. The AUC performances has improved by more than 0.003 going up to 0.017.

Table V.3 – Comparison of the proposed adaptation strategy

Database	Without biometric menagerie		With biometric menagerie	
	EER	AUC	EER	AUC
WEBGREYC	5.3%	0.02	0.8%	0.003
CMU	2.3%	0.004	0.3%	0.001

V.4 Seven categories user specific adaptive system

These experiments investigate an adaptive strategy that takes into account the specificities of each user to remedy to its intra-class-variations. More Doddington zoo categories are considered in these experiments.

V.4.1 Proposed adaptive strategy

Figure V.7 depicts the proposed authentication process based on the keystroke dynamics modality. Two samples are considered initially to register the typing manner of the user. Indeed, for recent password-based applications, users are usually asked to type their password and to confirm it when creating an account. Thus, since we can benefit from an additional capture for the creation of the reference, we considered two samples initially.

As the previous experiments, the classification is realized thanks to the GA-KNN classifier. During the two first update sessions, we start to classify users into two groups: sheep and goats. We are first interested to only these two groups because we focus on the most representative groups of the Doddington zoo.

Thereby, over the growing window phase, we assume that users, whose number of accepted queries has not overcome 3 samples during the update session, are not easily recognized. So, they are classified as goats. The rest of the users, those whose number of accepted queries is greater than 3, are classified as sheep, as they are easy to recognize.

For the sliding window mechanism, the size of the reference is no more significant as the maximum size of the reference is reached. So, we considered the Entropy measure to distinguish between the considered users groups. In fact, it was demonstrated in [Houmani & Garcia-Salicetti, 2016, Morales et al., 2014] that the higher the user's entropy is, the more the error rates increase. Thereby, both Personal and Relative Entropy are calculated according to equations (V.1) and (V.2) respectively.

Therefore, starting from session 4, we use the Entropy to classify users. We initially distinguish the lambs class. Once users of this class are defined, we determine during the

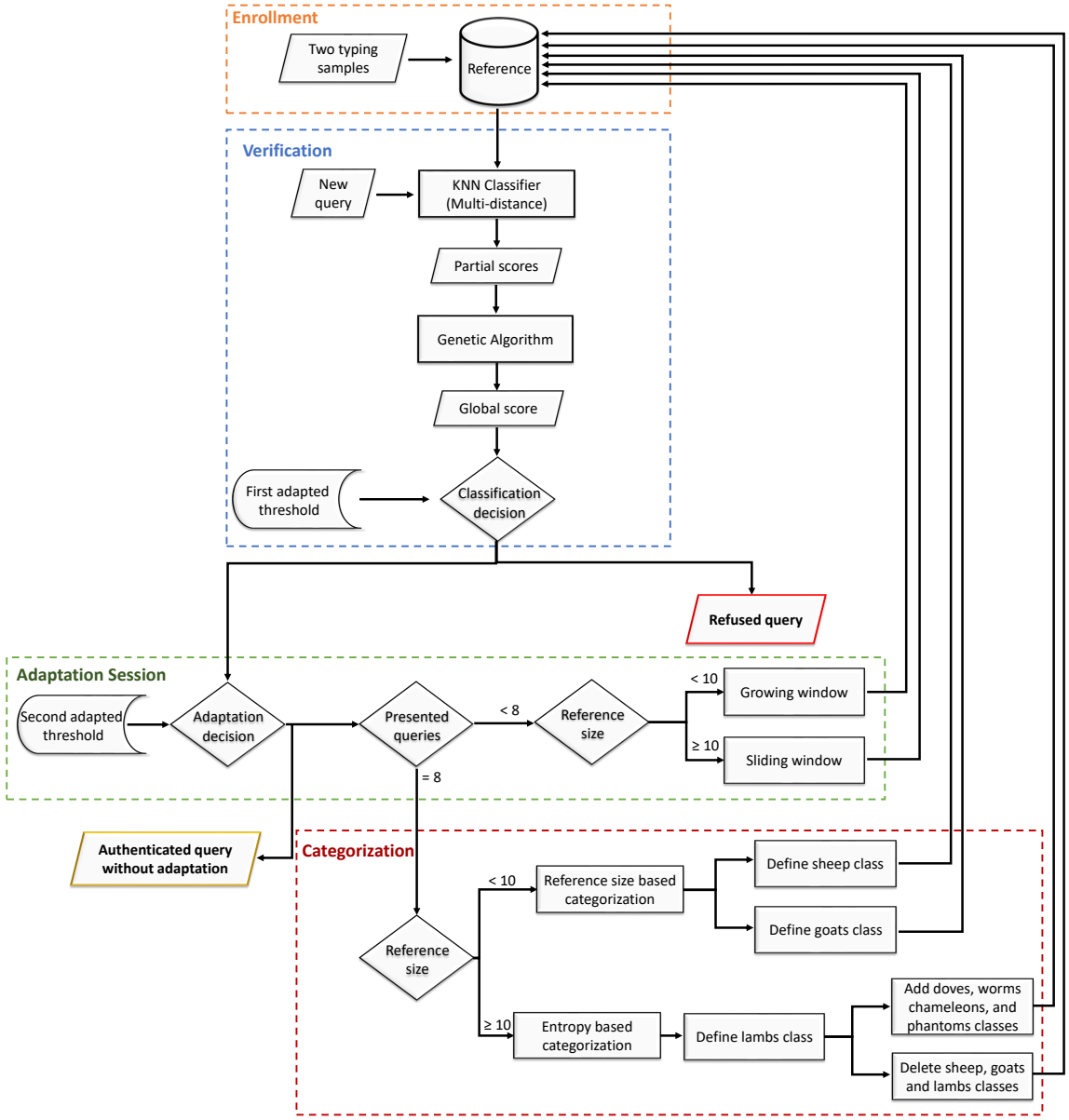


Figure V.7 – Description of the proposed keystroke authentication process.

following sessions the remaining classes of the zoo. Once session 6 starts, classes of worms, doves, chameleons and phantoms take place and classes of sheep, goats, and lambs disappear.

For each class, we use specific adaptation parameters [Mhenni et al., 2019a, Mhenni et al., 2018]. Concerning goats and worms classes, which are characterized by a high intra-class variations according to the different conducted experiments, we increased the maximum size of the reference to 15 in order to enrich the description of the keystroke dynamics of the users. The maximum size of phantoms class should be higher because this class is

difficult to describe. Regarding the lambs, worms, chameleons and phantoms classes, stricter thresholds are needed to minimize the acceptance of the impostor attacks. These thresholds are generated based on equation (V.4). The fixed parameters for each user category are detailed in Table V.4.

Table V.4 – Specific parameters according to user's category

User category	Reference size	Thresholds
Sheep	10	Adapted thresholds
Goats	15	Adapted thresholds
Lambs	10	Stricter thresholds
Worms	15	Stricter thresholds
Chameleons	10	Stricter thresholds
Doves	10	Adapted thresholds
Phantoms	20	Stricter thresholds

So, regarding users who suffer from a large intra-class variation, we enlarge the reference size to capture more variabilities. Moreover, for users that are more vulnerable to impostor attacks, we apply stricter thresholds to eliminate as much as possible the false accepted queries in our system.

V.4.2 Experiments and results

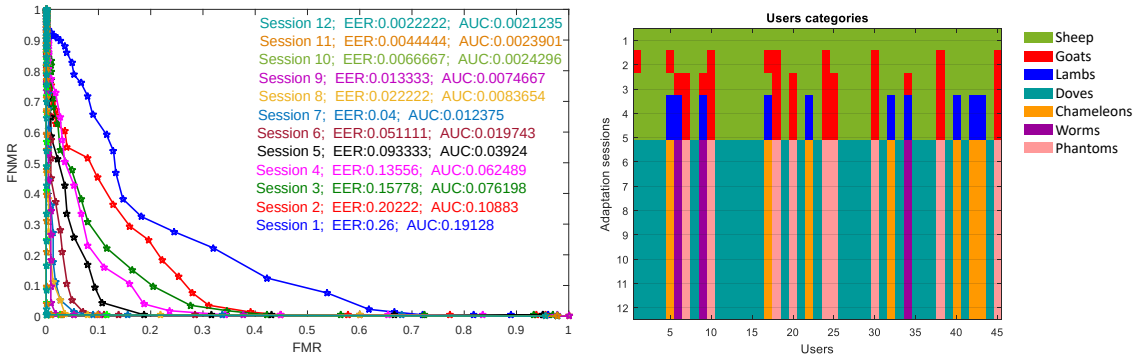
The proposed approach was tested on two public databases: WEBGREYC and CMU. We managed user samples during the adaptation sessions as follows. Two samples of each user are considered during the enrollment phase in order to create the reference. For each adaptation session, 8 new queries are introduced to the authentication system. These queries are divided into 5 genuine samples and 3 impostor ones. Thus, we considered 12 adaptation sessions for the WEBGREYC database and 80 adaptation sessions for the CMU database.

To evaluate the proposed approach we analyzed different data stream for each adaptation session:

- Scenario 1: Presenting 5 genuine samples first, afterwards 3 imposter samples are presented to the authentication system.
- Scenario 2: Presenting alternated genuine and imposter samples.
- Scenario 3: Presenting 3 imposter samples first, afterwards 5 genuine samples are presented to the authentication system.

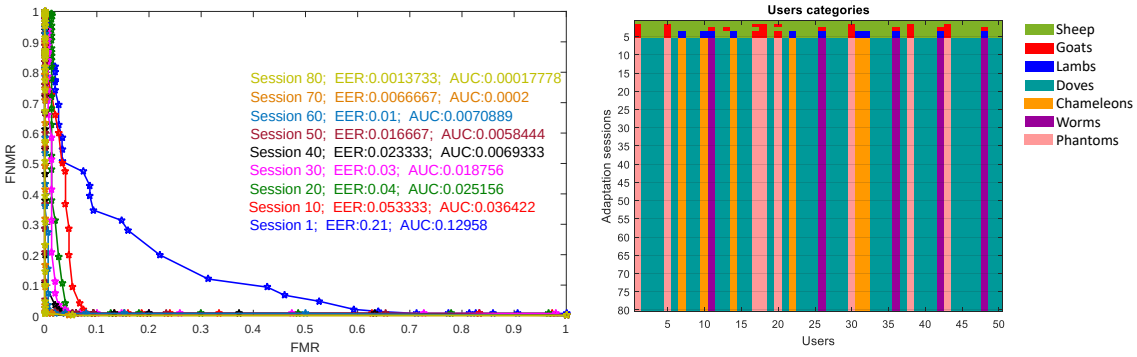
Generally, the first two data streams conveniently fit the actual scenarios of the password based applications [Mhenni et al., 2019a]. In fact, just after creating an account, the user is usually asked to enter his credentials again to gain access to his account. Consequently, at least one genuine query is guaranteed in the beginning of the process.

To evaluate the performance of the proposed approach, we consider two evaluation metrics: the EER and the AUC. The obtained results show an improvement in the performance of the strategy as demonstrated in Figures V.8(a) and V.9(a). Adding doves, phantoms, chameleons and worms classes, improved the EER performances by 0.6% for the WEBGREYC database and by 0.2% for the CMU database. Furthermore, when compared to the same adaptation approach without biometric menagerie, the user specific adaptation approach ensures an improved EER performance of more then 2% for CMU database and 5% for WEBGREYC database.



(a) DET curves. (b) Distribution of user classes.

Figure V.8 – Obtained performances and the distribution of users classes when considering scenario 1 for WEBGREYC database.



(a) DET curves. (b) Distribution of user classes.

Figure V.9 – Achieved performances and the distribution of users classes when considering scenario 1 for CMU database.

We also depict the distribution of users categories among all adaptation sessions for the two considered databases. The sheep class includes the majority of users as shown in Figures V.8 and V.9.

To illustrate the benefits of the consideration of 7 classes of the Doddington zoo in the proposed approach, we compared it to the same adaptation approach without biometric menagerie and with the consideration of only 3 classes conducted in [Mhenni et al., 2018a] namely sheep, goats and lambs. As demonstrated in Tables V.5 and V.6, the proposed approach show improved performances as it proposes an adaptive strategy that is the most appropriate to the user's specificities. In fact, the considered users' categories encompass a wider variety of users. Hence, the adaptation method acts according to each user's particularities.

Table V.5 – Comparison of the proposed adaptation strategy for WEBGREYC database

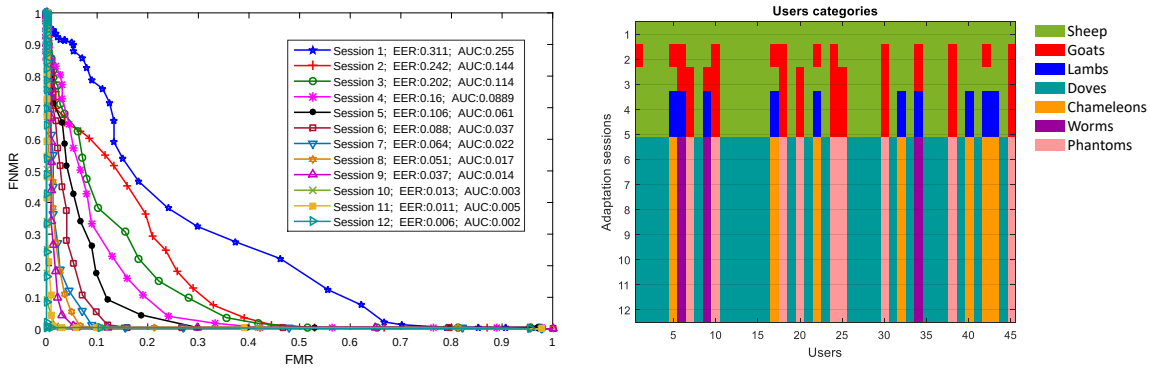
Adaptation strategy	EER	AUC
Without Doddington menagerie [Mhenni et al., 2019b]	5.3%	0.02
Biometric menagerie based on 3 classes [Mhenni et al., 2018a]	0.8 %	0.003
Biometric menagerie based on 7 classes	0.2%	0.002

Table V.6 – Comparison of the proposed adaptation strategy for CMU database

Adaptation strategy	EER	AUC
Without Doddington menagerie [Mhenni et al., 2019b]	2.3%	0.004
Biometric menagerie based on 3 classes [Mhenni et al., 2018a]	0.3%	0.001
Biometric menagerie based on 7 classes	0.1%	0.0001

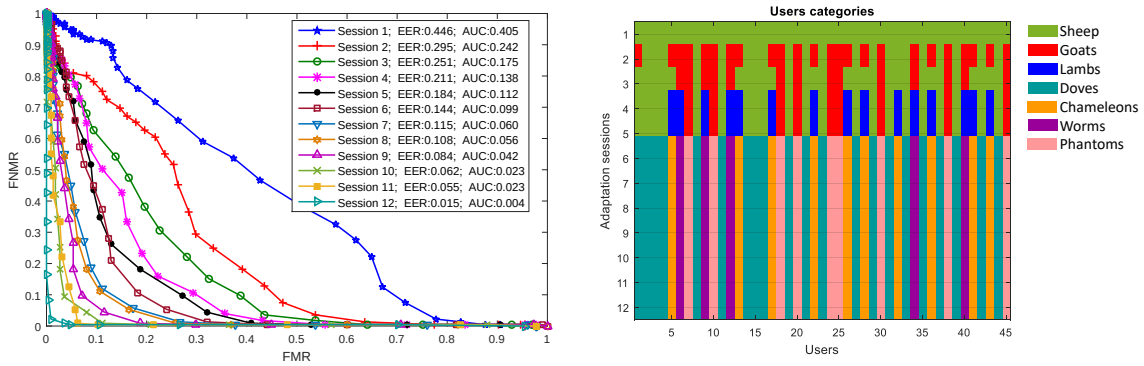
To reveal the impact of imposter attacks on the proposed approach, we tested different scenarios of the queries presentations. When considering the 3 imposter samples before the genuine ones (scenario 3), the performances are considerably decreased as demonstrated in Figure V.11(a). This is quite expected as the initial reference doesn't contain enough intra-class variation. Thus the recognition errors are higher in the beginning of the process. These errors decrease during the adaptation sessions through the proposed method. In addition, we illustrated the the users categorization in Figure V.11. It is quite clear that the number of users belonging to goats class has increased considerably since the beginning. In fact, the percentage of goats class in adaptation session 2 raised from 20% (for scenario 1) to 53% (for scenario 2). This may be due to the inclusion of some imposter samples in the reference. Subsequently, these imposter samples will be removed as time elapses due to the proposed

adaptation system. In fact, for scenario 3, the percentage of users associated to goats class in adaptation session 2 decreased to 31%. Thanks to the considered user specific parameters, the number of genuine samples included in the reference increase and the imposter samples decrease especially through the sliding window mechanism. Thus, the intra-class variation of the reference samples is reduced.



(a) DET curves. (b) Distribution of user classes.

Figure V.10 – Obtained performances and the distribution of users classes when considering scenario 2 for WEBGREYC database.



(a) DET curves. (b) Distribution of user classes.

Figure V.11 – Obtained performances and the distribution of users classes when considering scenario 3 for WEBGREYC database.

When mixing the genuine and imposter queries (scenario 2), the obtained results are better than those obtained in scenario 3 as depicted in Figure V.10(a) and they are quite similar to those of scenario 1. The scenarios presenting better performances (1 and 2) are more realistic. In fact, adding a new account is usually transparent for any password based application. Hence, it is not evident that a hacker encounters an account since its creation.

We have furthermore performed an analysis on the variation of the reference size concerning each user category of the menagerie considered in the proposed approach. As depicted in Tables V.7 and V.8, the chosen parameters are optimal. Indeed, minimizing the size of the reference, guaranteed the gain in used memory space, but no improvement in the performance is recorded. Moreover, while enlarging the size of the reference, a small increase in performance is registered. Thus, the extra memory space allocated does not produce a significant influence on the obtained performance. Hence, we prove that the chosen reference sizes are the most appropriate to each user category.

Table V.7 – Obtained performances by varying the size of the reference for each user category for WEBGREYC database

Reference size	**User category**							**Performances**	
	Sheep	Goats	Lambs	Worms	Chameleons	Doves	Phantoms	EER	AUC
max1	5	10	5	10	5	5	15	6.5%	0.05
max2	10	15	10	15	10	10	20	2.22%	0.002
max3	15	20	15	20	15	15	25	2%	0.0017

Table V.8 – Obtained performances by varying the size of the reference for each user category for CMU database

Reference size	**User category**							**Performances**	
	Sheep	Goats	Lambs	Worms	Chameleons	Doves	Phantoms	EER	AUC
max1	5	10	5	10	5	5	15	5.9%	0.047
max2	10	15	10	15	10	10	20	1.37%	0.0001
max3	15	20	15	20	15	15	25	1.14%	0.0008

V.5 Conclusion

In this chapter, we put forward a novel authentication method that helps to reinforce the security of IT services. Based on keystroke dynamics modality, the proposed method helps password based application to overcome hacking attacks. Indeed, the adaptive strategy specific to the user's category presents many advantages. First, the recognition of the user category according to the animal based categories of the Doddington Zoo, helps to distinguish the user's specificities. Then an adaptive strategy that remedy the problems of the user class is adopted. So, three different adaptive mechanism are simultaneously used : the growing window mechanism, the sliding window mechanism and the least frequently used mechanism.

Another important benefit of the proposed method is the minimization of the size of the reference. As it is user dependent, a gain in used memory is ensured. Only users with a large intra-class variation , have a larger reference size. Moreover, users who are more vulnerable to hacker attacks, are given stricter decision thresholds. Even if this choice minimizes the capture of intra-class variation of these users, since only the most similar data are considered, it protects them against attacks which are their weak point.

CHAPTER VI

General conclusion and future work

VI.1 General conclusion

Password-based applications have an important role in securing sensitive data and are ubiquitous in our daily lives. However, they are still vulnerable to hacker attacks. Thus, improving the security of these applications by adding keystroke dynamics verification remains a major challenge, as long as an accurate description of the user's typing manner is crucial to address the problem of intra-class variation.

Biometric systems are able to authenticate the identity of an individual based on what he is/does. They have been successfully used in several applications. Characteristics used for recognition should meet some properties [Jain et al., 2004a], as discussed in the beginning of this paper: universality, distinctiveness, permanence and collectability. However, recent studies have shown that the permanence is not met for several biometric modalities [Roli et al., 2007, Poh et al., 2012, Rattani, 2015] especially for keystroke dynamics modality [Giot et al., 2012c, Pisani et al., 2017]. This is due to several reasons, including ageing and changing conditions, as discussed in Section II.3.6. In order to deal with this problem, adaptive biometric systems have been proposed. This is relatively new field of study in biometrics.

This thesis work is part of the field of biometric reference adaptation. We proposed a contribution to the reference modeling by using a single sample for the enrollment phase, which is then enriched by the update strategy to enlarge the gallery size. This solution meets the requirements of today's applications especially those integrated in the web or mobile. We also proposed a new update decision criterion. This criterion guarantees individual and adaptive thresholds that are adequate to the user's typing rhythm. Also, the proposed adaptation mechanism "double serial" allows the enrichment and adaptation of the reference over time while guaranteeing a reduced size. This will indirectly influence user recognition performance. Finally, we proposed an update strategy specific to each user category according to Doddington zoo theory. We have considered increasingly a larger number of categories of the zoo to characterize the user precisely.

This work was conducted as follows:

In the first chapter, we presented the economic and cybernetic contexts as well as the motivations of our study on password based security. An overview on the different industrialized solutions to reinforce the applications security was presented.

In the second chapter, we have spread a review of the operating principle of biometric systems, especially for the keystroke dynamics modality. In addition, we presented the limitations of this modality and the challenges of its industrialization including the intra-class

variation of user characteristics as time elapses and the tedious enrollment phase. These limitations and variations that motivate the use of adaptation strategies.

The third chapter presented the state of the art on the different adaptation methods existing in the literature. We have also detailed adaptive strategies according to a new taxonomy to easily compare them. Regarding the proposed taxonomy, the update strategy were divided into five components: reference modeling, decision criterion, adaptation mode, periodicity of adaptation, and adaptation mechanism. We were able to make contributions on some parameters with remarkable enhancements to remedy to target disadvantages.

The fourth chapter investigates a novel method, which considers the conditions necessary for the application in real life of the keystroke dynamics modality especially for web services and mobile applications. In fact, in spite of its great advantage to reinforce the security of the password-based applications facing hacking attacks, this modality has not been industrialized yet. The main interest of the proposed method is that it minimizes as much as possible the number of samples used in the learning phase. Indeed, a unique sample is required initially. Besides, we adopt the double serial adaptation mechanism to remedy to the intra-class variations of the users' characteristics: It consists in combining the growing window and the sliding window mechanisms. The growing window serves to enlarge the users' galleries so as to capture more intra-class variability. After reaching the maximum size of the reference, which is fixed to 10, the sliding window mechanism takes place. It permits describing and following the temporal variation of the users' keystroke dynamics. Also, the adaptive threshold criterion has a great impact on the improvement of the obtained results. It is adapted to the gallery variation of each user. Eventually, the classification is achieved thanks the the GA-KNN classifier which was efficient especially since we eliminated the enrollment phase. Thanks to all these choices, we have obtained a competitive performance with a minimal size of the reference template (one sample for the enrollment and ten for the maximum size of the reference gallery). The accomplished results have been interesting.

In the fifth chapter, we put forward a user specific adaptation strategy based on the Doddington Zoo concept. It consists in applying an adaptive strategy related to each user class. So, regarding users who suffer from a large intra-class variation, we enlarge the reference size to capture more variabilities. Moreover, for users that are more vulnerable to impostor attacks, we apply stricter thresholds to eliminate as much as possible the false accepted queries in our system. Indeed, Doddington zoo is a biometric menagerie that applies an analogy between users and animals characteristics, and it was efficient for discrimination between users. A large number of the zoo classes is considered in this work, thus demonstrating enhanced performances. Besides, the proposed approach has the advantage of being conform to the web and mobile applications that generally consider only two password acquisitions (the

second is to confirm the firts typed one) when creating a new account. So, we consider only these two samples to create the user's reference.

VI.2 Future work

As perspectives, we are involved in a novel approach that may improve the performances of the first sessions so as to make the keystroke dynamics modality more compatible with industrialization conditions.

Furthermore, we aim to apply and model impostor attacks to reinforce the security of our authentication system. Thus, the possibility of extending the Doddington zoo classes by incorporating the wolves class to the considered ones and examine their relevance to the proposed system.

Besides, it will be worth applying and comparing the proposed method to other devices like mobile phones and to other modalities like voice and touch screen interactions. One possible orientation for typing dynamics is its use in touch screen devices due to their increasing availability. These devices can provide additional features to increase accuracy. As a result, more public databases on keystroke dynamics are needed. In addition, the use of more databases would increase confidence in performance comparisons of classifiers established in the literature.

An other opportunity consists in the acquisition of a novel dataset suitable to evaluate adaptive biometric systems. It can be oriented for mobile devices or for the keystroke dynamic of Arabic passwords (which is in progress). This dataset should meet some requirements to be used to evaluate adaptive biometric systems (see Section II.3.3). These datasets need to contain several samples per user and, ideally, they should be acquired at different acquisition sessions. Currently, there are some public datasets, however, additional ones with a greater number of users and sessions are still needed. One example is for the evaluation of large scale adaptive systems, as discussed in the previous item. Since it requires a lot of effort to be able to acquire biometric data for the same users during long periods, it is considered a current challenge in the field. The design of cohort databases is another related topic. These databases could be useful to implement score normalization [Poh et al., 2009a] for adaptive biometric systems.

In addition, we can use the user's profile for the update of the biometric reference. It can be used as an adaptation criterion to help when to apply the adaptation mechanism. Or even, it can be an important information to define the user's specificity and to set appropriate adaptation parameters. we can also define an attack strategy specific to the adaptation strategies. It may be a standard evaluation of update systems.

Moreover, deep learning opens promising new horizons. Exploiting this classifier for the keystroke dynamics modality is an interesting opportunity, taking into account the results already obtained and the choices already made.

References

[Ahmed & Traore, 2014]
Ahmed, A. A. & Traore, I. (2014). Biometric recognition based on free-text keystroke dynamics. *IEEE Transactions on Cybernetics*, 44(4), 458–472. (cited in 36).

[Akhtar et al., 2014]
Akhtar, Z., Ahmed, A., Erdem, C. E., & Foresti, G. L. (2014). Biometric template update under facial aging. In *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2014 IEEE Symposium on* (pp. 9–15).: IEEE. (cited in 56 and 81).

[Aljohani et al., 2018]
Aljohani, O., Aljohani, N., Bours, P., & Alsolami, F. (2018). Continuous authentication on pcs using artificial immune system. In *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)* (pp. 1–6).: IEEE. (cited in 28).

[Alsultan & Warwick, 2013]
Alsultan, A. & Warwick, K. (2013). User-friendly free-text keystroke dynamics authentication for practical applications. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* (pp. 4658–4663).: IEEE. (cited in 8).

[Anagun, 2006]
Anagun, A. (2006). Development of committee neural network for computer access security system. *Computational Science and Its Applications-ICCSA 2006*, (pp. 11–20). (cited in 36).

[Araújo et al., 2005]
Araújo, L. C., Sucupira, L. H., Lizarraga, M. G., Ling, L. L., & Yabu-Uti, J. B. T. (2005). User authentication through typing biometrics features. *Signal Processing, IEEE Transactions on*, 53(2), 851–855. (cited in 27).

[Bailly-Bailliére et al., 2003]
Bailly-Bailliére, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Mariéthoz, J., Matas, J., Messer, K., Popovici, V., Porée, F., Ruiz, B., & Thiran, J.-P. (2003). The banca database and evaluation protocol. In *Proceedings of the 4th international conference on Audio- and video-based biometric person authentication*, AVBPA'03 (pp. 625–638). Berlin, Heidelberg: Springer-Verlag. (cited in 69).

[Bartlow & Cukic, 2006]
Bartlow, N. & Cukic, B. (2006). Evaluating the reliability of credential hardening through keystroke dynamics. In *Software Reliability Engineering, 2006. ISSRE'06. 17th International Symposium on* (pp. 117–126).: IEEE. (cited in 30, 31, and 35).

[Bergadano et al., 2003]
Bergadano, F., Gunetti, D., & Picardi, C. (2003). Identity verification through dynamic keystroke analysis. *Intelligent Data Analysis*, 7(5), 469–496. (cited in 8).

[Bhargav-Spantzel et al., 2007]
Bhargav-Spantzel, A., Squicciarini, A. C., Modi, S., Young, M., Bertino, E., & Elliott, S. J. (2007). Privacy preserving multi-factor authentication with biometrics. *Journal of Computer Security*, 15(5), 529–560. (cited in viii and 22).

[Biggio et al., 2012]
Biggio, B., Fumera, G., Roli, F., & Didaci, L. (2012). Poisoning adaptive biometric systems. In *Proceedings of the 2012 Joint IAPR international conference on Structural, Syntactic, and Statistical Pattern Recognition*, SSPR'12/SPR'12 (pp. 417–425). Berlin, Heidelberg: Springer-Verlag. (cited in 45, 52, 55, 69, 70, and 72).

[Biggio et al., 2015]
Biggio, B., g. fumera, Russu, P., Didaci, L., & Roli, F. (2015). Adversarial biometric recognition : A review on biometric system security from the adversarial machine-learning perspective. *IEEE Signal Processing Magazine*, 32(5), 31–41. (cited in 52).

[Bishop, 2006]
Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. (cited in 48).

[Bleha et al., 1990a]
Bleha, S., Slivinsky, C., & Hussien, B. (1990a). Computer-access security systems using keystroke dynamics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(12), 1217–1222. (cited in 34 and 36).

[Bleha et al., 1990b]
Bleha, S., Slivinsky, C., & Hussien, B. (1990b). Computer-access security systems using keystroke dynamics. *IEEE Transactions on pattern analysis and machine intelligence*, 12(12), 1217–1222. (cited in 35).

[Bleha & Obaidat, 1993]
Bleha, S. A. & Obaidat, M. S. (1993). Computer users verification using the perceptron algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 900–902. (cited in 36).

[Blum & Chawla, 2001]
Blum, A. & Chawla, S. (2001). Learning from labeled and unlabeled data using graph mincuts. In *Proceedings of the Eighteenth international Conference on Machine Learning*. (cited in 57).

[Blum & Mitchell, 1998]
Blum, A. & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, COLT' 98 (pp. 92–100).: ACM. (cited in 53 and 63).

[Boechat et al., 2007]
Boechat, G. C., Ferreira, J. C., & Carvalho Filho, E. C. (2007). Authentication personal. In *Intelligent and Advanced Systems, 2007. ICIAS 2007. International Conference on* (pp. 254–256).: IEEE. (cited in 34 and 35).

[Bonneau et al., 2012]
Bonneau, J., Herley, C., Van Oorschot, P. C., & Stajano, F. (2012). The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *Security and Privacy (SP), 2012 IEEE Symposium on* (pp. 553–567).: IEEE. (cited in 11).

[Bours & Barghouthi, 2009]
Bours, P. & Barghouthi, H. (2009). Continuous authentication using biometric keystroke dynamics. In *The Norwegian Information Security Conference (NISK)*, volume 2009. (cited in 28).

[Bours & Ellingsen, 2018]
Bours, P. & Ellingsen, J. (2018). Cross keyboard keystroke dynamics. In *2018 1st International Conference on Computer Applications & Information Security (ICCAIS)* (pp. 1–6).: IEEE. (cited in 30 and 53).

[Bours & Mondal, 2015]
Bours, P. & Mondal, S. (2015). Continuous authentication with keystroke dynamics. *Norwegian Information Security Laboratory NISlab*, (pp. 41–58). (cited in 28 and 38).

[Brand, 1985]
Brand, S. (1985). Department of defense password management guideline. (cited in 12 and 14).

[Brown & Rogers, 1993]
Brown, M. & Rogers, S. J. (1993). User identification via keystroke characteristics of typed names using neural networks. *International Journal of Man-Machine Studies*, 39(6), 999–1014. (cited in 36).

[Buschek, 2018]
Buschek, D. (2018). *Behaviour-aware mobile touch interfaces*. PhD thesis, lmu. (cited in 36).

[Carls, 2009]
Carls, J. W. (2009). *A framework for analyzing biometric template aging and renewal prediction*. ProQuest. (cited in 49, 50, and 51).

[Çeker & Upadhyaya, 2016]
Çeker, H. & Upadhyaya, S. (2016). Adaptive techniques for intra-user variability in keystroke dynamics. (pp. 1–6). (cited in 60, 62, and 63).

[Ceker & Upadhyaya, 2016]
Ceker, H. & Upadhyaya, S. (2016). Adaptive techniques for intra-user variability in keystroke dynamics. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)* (pp. 1–6). (cited in 75).

[Chang et al., 2012]
Chang, T.-Y., Tsai, C.-J., & Lin, J.-H. (2012). A graphical-based password keystroke dynamic authentication system for touch screen handheld mobile devices. *Journal of Systems and Software*, 85(5), 1157–1165. (cited in 29, 30, and 35).

[Chang, 2006]
Chang, W. (2006). Reliable keystroke biometric system based on a small number of keystroke samples. In *Emerging Trends in Information and Communication Security* (pp. 312–320). Springer. (cited in 30, 34, and 35).

[Cheswick, 2013]
Cheswick, W. (2013). Rethinking passwords. *Communications of the ACM*, 56(2), 40–44. (cited in 14).

[Das et al., 2014]
Das, A., Bonneau, J., Caesar, M., Borisov, N., & Wang, X. (2014). The tangled web of password reuse. In *NDSS*, volume 14 (pp. 23–26). (cited in 14).

[Didaci et al., 2014]
Didaci, L., Marcialis, G. L., & Roli, F. (2014). Analysis of unsupervised template update in biometric recognition systems. *Pattern Recognition Letters*, 37, 151–160. (cited in 45 and 109).

[Doddington et al., 1998]
Doddington, G., Liggett, W., Martin, A., Przybocki, M., & Reynolds, D. (1998). *Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation*. Technical report, NATIONAL INST OF STANDARDS AND TECHNOLOGY GAITHERSBURG MD. (cited in 109).

[Drygajlo et al., 2009]
Drygajlo, A., Li, W., & Zhu, K. (2009). Q-stack aging model for face verification. In *Signal Processing Conference, 2009 17th European* (pp. 65–69).: IEEE. (cited in 88).

[Çeker & Upadhyaya, 2017]
Çeker, H. & Upadhyaya, S. (2017). Transfer learning in long-text keystroke dynamics. In *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)* (pp. 1–6). (cited in 60, 62, and 75).

[El Gayar et al., 2006]
El Gayar, N., Shaban, S. A., & Hamdy, S. (2006). Face recognition with semi-supervised learning and multiple classifiers. In *Proceedings of the 5th WSEAS International Conference on Computational Intelligence, Man-Machine Systems and Cybernetics, Venice, Italy* (pp. 296–301).: Citeseer. (cited in 65).

[El Kissi Ghalleb, 2017]
El Kissi Ghalleb, A. (2017). *Contribution á la reconnaissance des individus par fusion de modalités biométriques dures et douces du visage et du corps*. PhD thesis, Ecole Nationale D Ingénieurs de Monastir. (cited in 25).

[Elftmann, 2006]
Elftmann, P. (2006). Secure alternatives to password-based authentication mechanisms. *Lab. for Dependable Distributed Systems, RWTH Aachen Univ*. (cited in 29).

[Epp et al., 2011]
Epp, C., Lippold, M., & Mandryk, R. L. (2011). Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11 (pp. 715–724). New York, NY, USA: ACM. (cited in 109).

[Fernandez-Saavedra et al., 2016]
Fernandez-Saavedra, B., Sanchez-Reillo, R., Ros-Gomez, R., & Liu-Jimenez, J. (2016). Small fingerprint scanners used in mobile devices: the impact on biometric performance. *IET Biometrics*, 5(1), 28–36. (cited in 76).

[Florêncio & Herley, 2010]
Florêncio, D. & Herley, C. (2010). Where do security policies come from? In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10 (pp. 10:1–10:14). New York, NY, USA: ACM. (cited in 13 and 14).

[Florêncio et al., 2014]
Florêncio, D., Herley, C., & Van Oorschot, P. C. (2014). Password portfolios and the finite-effort user: Sustainably managing large numbers of accounts. In *USENIX Security Symposium* (pp. 575–590). (cited in 14).

[Freni et al., 2008a]
Freni, B., Marcialis, G. L., & Roli, F. (2008a). Replacement algorithms for fingerprint template update. In *Image Analysis and Recognition* (pp. 884–893). Springer. (cited in 52, 59, 61, and 62).

[Freni et al., 2008b]
Freni, B., Marcialis, G. L., & Roli, F. (2008b). Template selection by editing algorithms: A case study in face recognition. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (pp. 745–754).: Springer. (cited in 52, 66, and 67).

[Gaines et al., 1980]
Gaines, R. S., Lisowski, W., Press, S. J., & Shapiro, N. (1980). *Authentication by keystroke timing: Some preliminary results*. Technical report, DTIC Document. (cited in 22).

[Gate, 1972]
Gate, G. W. (1972). The reduced nearest neighbor rule. *IEEE Trans. Inf. Theory,*, 18(3), 431?433. (cited in 67).

[Giot, 2012]
Giot, R. (2012). *Contributions à la dynamique de frappe au clavier: multibiométrie, biométrie douce et mise à jour de la référence*. PhD thesis, Université de Caen. (cited in 26 and 88).

[Giot et al., 2011a]
Giot, R., Dorizzi, B., & Rosenberger, C. (2011a). Analysis of template update strategies for keystroke dynamics. In *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2011 IEEE Workshop on* (pp. 21–28).: IEEE. (cited in 45, 56, and 68).

[Giot et al., 2011b]
Giot, R., El-Abed, M., Hemery, B., & Rosenberger, C. (2011b). Unconstrained keystroke dynamics authentication with shared secret. *Computers & security*, 30(6), 427–445. (cited in 17, 22, 27, 30, 31, 33, 35, 36, 41, 52, 54, 75, 76, 81, 84, 85, 86, 87, 91, and 102).

[Giot et al., 2009a]
Giot, R., El-Abed, M., & Rosenberger, C. (2009a). Greyc keystroke: a benchmark for keystroke dynamics biometric systems. In *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009)* (pp. 419–424). Washington, District of Columbia, USA: IEEE Computer Society. (cited in 33 and 34).

[Giot et al., 2009b]
Giot, R., El-Abed, M., & Rosenberger, C. (2009b). Keystroke dynamics with low constraints svm based passphrase enrollment. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on* (pp. 1–6).: IEEE. (cited in 30, 31, 34, and 35).

[Giot et al., 2012a]
Giot, R., El-Abed, M., & Rosenberger, C. (2012a). Web-based benchmark for keystroke dynamics biometric systems: A statistical analysis. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on* (pp. 11–15).: IEEE. (cited in 30, 33, 34, 35, 56, 111, and 117).

[Giot et al., 2012b]
Giot, R., Rosenberger, C., & Dorizzi, B. (2012b). Can chronological information be used as a soft biometric in keystroke dynamics? In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on* (pp. 7 –10). (cited in 68).

[Giot et al., 2012c]
Giot, R., Rosenberger, C., & Dorizzi, B. (2012c). Hybrid template update system for unimodal biometric systems. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on* (pp. 1–7).: IEEE. (cited in xiii, 45, 48, 55, 63, 64, 65, 66, 68, 71, 72, 87, 95, and 130).

[Giot et al., 2012d]
Giot, R., Rosenberger, C., & Dorizzi, B. (2012d). Performance evaluation of biometric template update. In *International Biometric Performance Testing Conference (IBPC 2012)*. (cited in 67).

[Giot et al., 2013]
Giot, R., Rosenberger, C., & Dorizzi, B. (2013). A new protocol to evaluate the resistance of template update systems against zero-effort attacks. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on* (pp. 131–137). (cited in 68, 69, 71, and 72).

[Grabham & White, 2008]
Grabham, N. & White, N. (2008). Use of a novel keypad biometric for enhanced user identity verification. In *Instrumentation and Measurement Technology Conference Proceedings, 2008. IMTC 2008. IEEE* (pp. 12–16).: IEEE. (cited in 48 and 59).

[Grassi et al., 2017]
Grassi, P. A., Fenton, J. L., Newton, E. M., Perlner, R. A., Regenscheid, A. R., Burr, W. E., Richer, J. P., Lefkovitz, N. B., Danker, J. M., Choong, Y.-Y., Greene, K. K., & Theofanos, M. F. (2017). Digital identity guidelines authentication and lifecycle management. *NIST Special Publication 800-63B*. (cited in 15).

[Gunetti & Picardi, 2005]
Gunetti, D. & Picardi, C. (2005). Keystroke analysis of free text. *ACM Transactions on Information and System Security (TISSEC)*, 8(3), 312–347. (cited in 30, 34, and 35).

[Habib et al., 2017]
Habib, H., Colnago, J., Melicher, W., Ur, B., Segreti, S., Bauer, L., Christin, N., & Cranor, L. (2017). Password creation in the presence of blacklists. *Proc. USEC*, (pp.5̃0). (cited in 13).

[Hart, 1968]
Hart, P. E. (1968). The condensed nearest neighbor rule. *IEEE Trans. Inform. Theory (Corresp.)*, IT-14, 515–516,. (cited in 67).

[Herley & Van Oorschot, 2012]
Herley, C. & Van Oorschot, P. (2012). A research agenda acknowledging the persistence of passwords. *IEEE Security & Privacy*, 10(1), 28–36. (cited in 11).

[Himaga & Kou, 2008]
Himaga, M. & Kou, K. (2008). Finger vein authentication technology and financial applications. In N. Ratha & V. Govindaraju (Eds.), *Advances in Biometrics* (pp. 89–105). Springer London. (cited in 37).

[Hocquet, 2007]
Hocquet, S. (2007). *Authentification biométrique adaptative: application à la dynamique de frappe et à la signature manuscrite*. PhD thesis, Tours. (cited in 21).

[Hocquet et al., 2006]
Hocquet, S., Ramel, J.-Y., & Cardot, H. (2006). Estimation of user specific parameters in one-class problems. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4 (pp. 449–452).: IEEE. (cited in 88).

[Hocquet et al., 2007]
Hocquet, S., Ramel, J.-Y., & Cardot, H. (2007). User classification for keystroke dynamics authentication. In *International Conference on Biometrics* (pp. 531–539).: Springer. (cited in 27, 34, 82, and 83).

[Hosseinzadeh & Krishnan, 2008]
Hosseinzadeh, D. & Krishnan, S. (2008). Gaussian mixture modeling of keystroke patterns for biometric applications. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(6), 816–826. (cited in 30, 31, 35, and 88).

[Houmani & Garcia-Salicetti, 2016]
Houmani, N. & Garcia-Salicetti, S. (2016). On hunting animals of the biometric menagerie for online signature. *PloS one*, 11(4), e0151691. (cited in ix, 110, 111, and 121).

[Idrus et al., 2014]
Idrus, S. Z. S., Cherrier, E., Rosenberger, C., & Bours, P. (2014). Soft biometrics for keystroke dynamics: Profiling individuals while typing passwords. *Computers & Security*, 45, 147–155. (cited in 27, 30, and 41).

[Jain et al., 2004a]
Jain, A., Ross, A., & Prabhakar, S. (2004a). An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1), 4–20. (cited in 23, 43, and 130).

[Jain et al., 2004b]
Jain, A. K., Dass, S. C., & Nandakumar, K. (2004b). Soft biometric traits for personal recognition systems. In *Biometric Authentication* (pp. 731–738). Springer. (cited in 41).

[Jain et al., 2016]
Jain, A. K., Nandakumar, K., & Ross, A. (2016). 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, 79, 80 – 105. (cited in 40, 43, and 44).

[Jain et al., 2011]
Jain, A. K., Ross, A. A., & Nandakumar, K. (2011). *Introduction to biometrics*. Springer Science & Business Media. (cited in 23).

[Jiang & Ser, 2002]
Jiang, X. & Ser, W. (2002). Online fingerprint template improvement. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(8), 1121–1126. (cited in 53).

[Jindal et al., 2016]
Jindal, V., Birjandtalab, J., Pouyan, M. B., & Nourani, M. (2016). An adaptive deep learning approach for ppg-based identification. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the* (pp. 6401–6404).: IEEE. (cited in 26).

[Jobusch & Oldehoeft, 1989]
Jobusch, D. L. & Oldehoeft, A. E. (1989). A survey of password mechanisms: Weaknesses and potential improvements. part 1. *Computers & Security*, 8(7), 587–604. (cited in 11).

[Kacem et al., 2012]
Kacem, A., Aouïti, N., & Belaïd, A. (2012). Structural features extraction for handwritten arabic personal names recognition. In *2012 International Conference on Frontiers in Handwriting Recognition* (pp. 268–273).: IEEE. (cited in 25).

[Kacem & Saïdani, 2017]
Kacem, A. & Saïdani, A. (2017). A texture-based approach for word script and nature identification. *Pattern Analysis and Applications*, 20(4), 1157–1167. (cited in 25).

[Kang et al., 2007]
Kang, P., Hwang, S.-s., & Cho, S. (2007). Continual retraining of keystroke dynamics based authenticator. In *Advances in Biometrics* (pp. 1203–1211). Springer. (cited in xiii, 48, 50, 56, 59, 60, 62, and 68).

[Kekre & Bharadi, 2009]
Kekre, H. & Bharadi, V. (2009). Adaptive feature set updating algorithm for multimodal biometrics. In *Proceedings of the International Conference on Advances in Computing, Communication and Control* (pp. 277–282).: ACM. (cited in 49).

[Killourhy & Maxion, 2008]
Killourhy, K. & Maxion, R. (2008). The effect of clock resolution on keystroke dynamics. In *International Workshop on Recent Advances in Intrusion Detection* (pp. 331–350).: Springer. (cited in 30, 32, 34, 35, and 36).

[Killourhy & Maxion, 2010]
Killourhy, K. & Maxion, R. (2010). Why did my detector do that?! predicting keystroke-dynamics error rates. In S. Jha, R. Sommer, & C. Kreibich (Eds.), *Recent Advances in Intrusion Detection*, volume 6307 of *Lecture Notes in Computer Science* (pp. 256–276). Springer Berlin / Heidelberg. (cited in 27, 30, 33, 34, 35, and 117).

[Killourhy et al., 2009]
Killourhy, K. S., Maxion, R., et al. (2009). Comparing anomaly-detection algorithms for keystroke dynamics. In *Dependable Systems & Networks, 2009. DSN'09. IEEE/IFIP International Conference on* (pp. 125–134).: IEEE. (cited in 75).

[Killourhy & Maxion, 2012]
Killourhy, K. S. & Maxion, R. A. (2012). Free vs. transcribed text for keystroke-dynamics evaluations. In *Proceedings of the 2012 Workshop on Learning from Authoritative Security Experiment Results* (pp. 1–8).: ACM. (cited in 30 and 35).

[Komanduri et al., 2011]
Komanduri, S., Shay, R., Kelley, P. G., Mazurek, M. L., Bauer, L., Christin, N., Cranor, L. F., & Egelman, S. (2011). Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2595–2604).: ACM. (cited in 15).

[Li & Bours, 2018]
Li, G. & Bours, P. (2018). Studying wifi and accelerometer data based authentication method on mobile phones. In *Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications* (pp. 18–23).: ACM. (cited in 36).

[Lumini & Nanni, 2006]
Lumini, A. & Nanni, L. (2006). A clustering method for automatic biometric template selection. *Pattern Recognition*, 39(3), 495–497. (cited in 48).

[Magalhães et al., 2005]
Magalhães, S. T., Revett, K., & Santos, H. M. D. (2005). Password secured sites - stepping forward with keystroke dynamics. In *Proceedings of the International Conference on Next Generation Web Services Practices*, NWESP '05 (pp. 293–298).: IEEE Computer Society. (cited in 63 and 65).

[Marasco & Ross, 2015]
Marasco, E. & Ross, A. (2015). A survey on antispoofing schemes for fingerprint recognition systems. *ACM Computing Surveys (CSUR)*, 47(2), 28. (cited in 76).

[Masso & Vaisman, 2010]
Masso, M. & Vaisman, I. I. (2010). Accurate and efficient gp120 v3 loop structure based models for the determination of hiv-1 co-receptor usage. *BMC Bioinformatics*, 11(1), 1–11. (cited in 37).

[Messerman et al., 2011]
Messerman, A., Mustafić, T., Camtepe, S. A., & Albayrak, S. (2011). Continuous and non-intrusive identity verification in real-time environments based on free-text keystroke dynamics. In *Biometrics (IJCB), 2011 International Joint Conference on* (pp. 1–8).: IEEE. (cited in 28).

[Mhenni et al., 2016]
Mhenni, A., Cherrier, E., Rosenberger, C., & Essoukri Ben Amara, N. (2016). Keystroke template update with adapted thresholds. In *International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 483–488). (cited in 72, 88, 109, and 116).

[Mhenni et al., 2018a]
Mhenni, A., Cherrier, E., Rosenberger, C., & Essoukri Ben Amara, N. (2018a). Adaptive biometric strategy using doddington zoo classification of user's keystroke dynamics. In *2018 14th International Wireless Communications Mobile Computing Conference (IWCMC)* (pp. 488–493). (cited in 125).

[Mhenni et al., 2018b]
Mhenni, A., Cherrier, E., Rosenberger, C., & Essoukri Ben Amara, N. (2018b). Towards a secured authentication based on an online double serial adaptive mechanism of users' keystroke dynamics. In *International Conference on Digital Society and eGovernments (ICDS)* (pp. 73–80). (cited in 97 and 113).

[Mhenni et al., 2018]
Mhenni, A., Cherrier, E., Rosenberger, C., & Essoukri Ben Amara, N. (2018). User dependent template update for keystroke dynamics recognition. In *2018 International Conference on Cyberworlds (CW)* (pp. 324–330). (cited in 122).

[Mhenni et al., 2019a]
Mhenni, A., Cherrier, E., Rosenberger, C., & Essoukri Ben Amara, N. (2019a). Analysis of doddington zoo classification for user dependent template update: Application to keystroke dynamics recognition. *Future Generation Computer Systems*, 97, 210 – 218. (cited in 122 and 124).

[Mhenni et al., 2019b]
Mhenni, A., Cherrier, E., Rosenberger, C., & Essoukri Ben Amara, N. (2019b). Double serial adaptation mechanism for keystroke dynamics authentication based on a single password. *Computers & Security*, 83, 151 – 166. (cited in 113 and 125).

[Monaro et al., 2017]
Monaro, M., Spolaor, R., Li, Q., Conti, M., Gamberini, L., & Sartori, G. (2017). Type me the truth!: Detecting deceitful users via keystroke dynamics. In *Proceedings of the 12th International Conference on Availability, Reliability and Security* (pp.60).: ACM. (cited in 30 and 32).

[Mondal, 2016]
Mondal, S. (2016). *Continuous user authentication and identification: Combination of security & forensics*. PhD thesis. (cited in 26).

[Monrose & Rubin, 2000]
Monrose, F. & Rubin, A. D. (2000). Keystroke dynamics as a biometric for authentication. *Future Generation computer systems*, 16(4), 351–359. (cited in 30, 34, and 35).

[Montalvão et al., 2015]
Montalvão, J., Freire, E. O., Bezerra Jr., M. A., & Garcia, R. (2015). Contributions to empirical analysis of keystroke dynamics in passwords. *Pattern Recognition Letters*, 52(C), 80–86. (cited in 40).

[Montalvao et al., 2006]
Montalvao, J., Almeida, C. A. S., & Freire, E. O. (2006). Equalization of keystroke timing histograms for improved identification performance. In *Telecommunications Symposium, 2006 International* (pp. 560–565).: IEEE. (cited in 29, 30, 34, and 35).

[Montalvão Filho & Freire, 2006]
Montalvão Filho, J. R. & Freire, E. O. (2006). On the equalization of keystroke timing histograms. *Pattern Recognition Letters*, 27(13), 1440–1446. (cited in 30 and 35).

[Morales et al., 2014]
Morales, A., Fierrez, J., & Ortega-Garcia, J. (2014). Towards predicting good users for biometric recognition based on keystroke dynamics. In *European Conference on Computer Vision* (pp. 711–724).: Springer. (cited in 109 and 121).

[Morris & Thompson, 1979]
Morris, R. & Thompson, K. (1979). Password security: A case history. *Communications of the ACM*, 22(11), 594–597. (cited in 11).

[Muliono et al., 2018]
Muliono, Y., Ham, H., & Darmawan, D. (2018). Keystroke dynamic classification using machine learning for password authorization. *Procedia Computer Science*, 135, 564–569. (cited in 30 and 35).

[Nahin et al., 2014]
Nahin, A. N. H., Alam, J. M., Mahmud, H., & Hasan, K. (2014). Identifying emotion by keystroke dynamics and text pattern analysis. *Behaviour & Information Technology*, 33(9), 987–996. (cited in 109).

[Nilsen, 2018]
Nilsen, P.-K. (2018). Combining periodic and continuous authentication using keystroke dynamics. Master's thesis, NTNU. (cited in 28).

[Noval & López, 2008]
Noval, R. G. & López, F. P. (2008). Adaptative templates in biometric authentication. In *The 16th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, volume 2008 (pp.1̃4). (cited in 49, 50, and 51).

[Obaidat & Sadoun, 1997]
Obaidat, M. S. & Sadoun, B. (1997). Verification of computer users using keystroke dynamics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 27(2), 261–269. (cited in 75).

[Othman, 2016]
Othman, N. (2016). *Fusion techniques for iris recognition in degraded sequences*. PhD thesis, Paris Saclay. (cited in 25).

[Pagano et al., 2015]
Pagano, C., Granger, E., Sabourin, R., Tuveri, P., Marcialis, G., & Roli, F. (2015). Context-sensitive self-updating for adaptive face recognition. In *Adaptive Biometric Systems* (pp. 9–34). Springer. (cited in 49, 50, and 51).

[Passeri, 2017]
Passeri, P. (2017). Information security timelines and statistics. *hackmageddon.com*. (cited in 9).

[Pinto et al., 2014]
Pinto, P., Patrão, B., & Santos, H. (2014). Free typed text using keystroke dynamics for continuous authentication. In *IFIP International Conference on Communications and Multimedia Security* (pp. 33–45).: Springer. (cited in 28).

[Pisani, 2017]
Pisani, P. H. (2017). *Biometrics in a data stream context*. PhD thesis, Universidade de São Paulo (USP) - Instituto de Ciências Matemáticas e de Computação (ICMC). (cited in 26 and 41).

[Pisani et al., 2016]
Pisani, P. H., Giot, R., de Carvalho, A. C. P. L. F., & Lorena, A. C. (2016). Enhanced template update: Application to keystroke dynamics. *Computers & Security*, 60, 134–153. (cited in 17, 35, 36, 48, 49, 51, 52, 53, 65, 66, 68, 71, 72, 75, 81, 95, 101, and 102).

[Pisani et al., 2014]
Pisani, P. H., Lorena, A. C., & de Carvalho, A. C. P. L. F. (2014). Adaptive algorithms in accelerometer biometrics. In *2014 Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 336–341).: IEEE. (cited in 50, 51, 52, 60, and 62).

[Pisani et al., 2015a]
Pisani, P. H., Lorena, A. C., & de Carvalho, A. C. P. L. F. (2015a). Adaptive approches for keystroke dynamics. In *Neural Networks (IJCNN), The 2015 International Joint Conference on* (pp. 1–8). (cited in xiii, 50, 51, 52, 54, 60, 61, 62, 63, 65, 66, 68, 71, and 72).

[Pisani et al., 2015b]
Pisani, P. H., Lorena, A. C., & de Carvalho, A. C. P. L. F. (2015b). Adaptive positive selection for keystroke dynamics. *Journal of Intelligent & Robotic Systems*, 80(1), 277–293. (cited in 39, 48, 50, 51, 52, 60, 62, and 63).

[Pisani et al., 2015c]
Pisani, P. H., Lorena, A. C., & de Carvalho, A. C. P. L. F. (2015c). Ensemble of adaptive algorithms for keystroke dynamics. In *2015 Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 310–315).: IEEE. (cited in 65 and 66).

[Pisani et al., 2017]
Pisani, P. H., Poh, N., de Carvalho, A. C. P. L. F., & Lorena, A. C. (2017). Score normalization applied to adaptive biometric systems. *Computers & Security*, 70, 565 – 580. (cited in 30, 50, 51, 52, 68, 72, and 130).

[Poh et al., 2010a]
Poh, N., Kittler, J., & Bourlai, T. (2010a). Quality-based score normalization with device qualitative information for multimodal biometric fusion. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(3), 539–554. (cited in 40).

[Poh et al., 2015a]
Poh, N., Kittler, J., Chan, C.-H., & Pandit, M. (2015a). Algorithm to estimate biometric performance change over time. *IET Biometrics*, 4(4), 236–245. (cited in 39 and 109).

[Poh et al., 2010b]
Poh, N., Kittler, J., Marcel, S., Matrouf, D., & Bonastre, J.-F. (2010b). Model and score adaptation for biometric systems: Coping with device interoperability and changing acquisition conditions. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 1229–1232).: IEEE. (cited in 49, 50, 51, and 52).

[Poh et al., 2014]
Poh, N., Kittler, J., & Rattani, A. (2014). Handling session mismatch by fusion-based co-training: An empirical study using face and speech multimodal biometrics. In *2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM)* (pp. 81–86). (cited in 37, 64, 66, 69, and 70).

[Poh et al., 2015b]
Poh, N., Kittler, J., & Rattani, A. (2015b). Handling session mismatch by semi-supervised-based co-training scheme. In *Adaptive Biometric Systems* (pp. 35–49). Springer. (cited in 53 and 64).

[Poh et al., 2009a]
Poh, N., Merati, A., & Kittler, J. (2009a). Adaptive client-impostor centric score

normalization: A case study in fingerprint verification. In *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems* (pp. 1–7). (cited in 50, 52, and 132).

[Poh et al., 2012]
Poh, N., Rattani, A., & Roli, F. (2012). Critical analysis of adaptive biometric systems. *IET biometrics*, 1(4), 179–187. (cited in 43, 45, 68, 109, and 130).

[Poh et al., 2009b]
Poh, N., Wong, R., Kittler, J., & Roli, F. (2009b). Challenges and research directions for adaptive biometric recognition systems. In *Advances in Biometrics* (pp. 753–764). Springer. (cited in 40, 45, and 70).

[Precise Biometrics, 2014]
Precise Biometrics (2014). Understanding biometric performance evaluation. (cited in 37).

[Radha et al., 2016]
Radha, R., Blesswin, A. J., & Mary, G. S. (2016). A simple innovative approach dna-based saliva security system for user authentication. *Indian Journal of Science and Technology*, 9(37). (cited in 26).

[Rattani, 2010]
Rattani, A. (2010). Adaptive biometric system based on template update procedures. *Dept. of Elect. and Comp. Eng., University of Cagliari, PhD Thesis*. (cited in 41, 49, 50, 51, 56, and 87).

[Rattani, 2015]
Rattani, A. (2015). *Introduction to Adaptive Biometric Systems*, (pp. 1–8). Springer International Publishing. (cited in 130).

[Rattani et al., 2007]
Rattani, A., Kisku, D., Lagorio, A., & Tistarelli, M. (2007). Facial template synthesis based on sift features. In *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on* (pp. 69–73).: IEEE. (cited in 80).

[Rattani et al., 2012]
Rattani, A., Marcialis, G. L., Granger, E., & Roli, F. (2012). A dual-staged classification-selection approach for automated update of biometric templates. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 2972–2975).: IEEE. (cited in 57 and 58).

[Rattani et al., 2008a]
Rattani, A., Marcialis, G. L., & Roli, F. (2008a). Biometric template update using the graph mincut algorithm: A case study in face verification. In *Biometrics Symposium, 2008. BSYM'08* (pp. 23–28).: IEEE. (cited in xiii, 45, 55, 57, and 58).

[Rattani et al., 2008b]
Rattani, A., Marcialis, G. L., & Roli, F. (2008b). Capturing large intra-class variations of biometric data by template co-updating. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on* (pp. 1–6).: IEEE. (cited in 63, 65, and 66).

[Rattani et al., 2009]
Rattani, A., Marcialis, G. L., & Roli, F. (2009). *An Experimental Analysis of the Relationship between Biometric Template Update and the Doddington's Zoo: A Case Study in Face Verification*, (pp. 434–442). Springer Berlin Heidelberg. (cited in 39).

[Rattani et al., 2011]
Rattani, A., Marcialis, G. L., & Roli, F. (2011). Self adaptive systems: An experimental analysis of the performance over time. In *Computational Intelligence in Biometrics and Identity Management (CIBIM), 2011 IEEE Workshop on* (pp. 36–43).: IEEE. (cited in 38, 53, 56, and 88).

[Rattani et al., 2013a]
Rattani, A., Marcialis, G. L., & Roli, F. (2013a). Biometric system adaptation by self-update and graph-based techniques. *Journal of Visual Languages & Computing*, 24(1), 1 – 9. (cited in xiii, 57, 58, 70, and 72).

[Rattani et al., 2013b]
Rattani, A., Marcialis, G. L., & Roli, F. (2013b). Biometric system adaptation by self-update and graph-based techniques. *Journal of Visual Languages & Computing*, 24(1), 1–9. (cited in 56, 57, and 67).

[Rattani et al., 2013c]
Rattani, A., Marcialis, G. L., & Roli, F. (2013c). A multi-modal dataset, protocol and tools for adaptive biometric systems: a benchmarking study. *IJBM*, 5(3/4), 266–287. (cited in xiii, 56, 63, 64, 65, 66, 68, 70, 71, and 72).

[Revett et al., 2006]
Revett, K., De Magalhães, S. T., & Santos, H. M. (2006). Enhancing login security through the use of keystroke input dynamics. In *International Conference on Biometrics* (pp. 661–667).: Springer. (cited in 22 and 34).

[Reynolds & Rose, 1995]
Reynolds, D. A. & Rose, R. C. (1995). Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE transactions on speech and audio processing*, 3(1), 72–83. (cited in 48).

[Ritter et al., 1975]
Ritter, G., Woodruff, H., Lowry, S., & Isenhour, T. (1975). An algorithm for a selective nearest neighbor decision rule. *IEEE Transactions on Information Theory*, 21(6), 665–669. (cited in 67).

[Rodrigues et al., 2005]
Rodrigues, R., Yared, G., do N. Costa, C., Yabu-Uti, J., Violaro, F., & Ling, L. (2005). Biometric access control through numerical keyboards based on keystroke dynamics. *Advances in biometrics*, (pp. 640–646). (cited in 36).

[Rodrigues et al., 2006]
Rodrigues, R. N., Yared, G. F., Costa, C. R. d. N., Yabu-Uti, J. B., Violaro, F., & Ling, L. L. (2006). Biometric access control through numerical keyboards based on keystroke dynamics. In *International Conference on Biometrics* (pp. 640–646).: Springer. (cited in 30 and 35).

[Roli et al., 2008]
Roli, F., Didaci, L., & Marcialis, G. (2008). Adaptive biometric systems that can improve with use. In N. Ratha & V. Govindaraju (Eds.), *Advances in Biometrics* (pp. 447–471). Springer London. (cited in 43 and 56).

[Roli et al., 2007]
Roli, F., Didaci, L., & Marcialis, G. L. (2007). Template co-update in multimodal biometric systems. In *Advances in Biometrics* (pp. 1194–1202). Springer. (cited in 49, 51, 52, 63, 65, 66, 68, and 130).

[Roli & Marcialis, 2006]
Roli, F. & Marcialis, G. L. (2006). Semi-supervised pca-based face recognition using self-training. In *Structural, Syntactic, and Statistical Pattern Recognition* (pp. 560–568). Springer. (cited in xiii, 45, 56, 57, 58, 67, and 68).

[Ross et al., 2009]
Ross, A., Rattani, A., & Tistarelli, M. (2009). Exploiting the doddington zoo effect in biometric fusion. In *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, (BTAS)* (pp. 1–7).: IEEE. (cited in 109).

[Ross et al., 2006]
Ross, A. A., Nandakumar, K., & Jain, A. K. (2006). *Handbook of multibiometrics*, volume 6. Springer Science & Business Media. (cited in 40).

[Rybnicek et al., 2014]
Rybnicek, M., Lang-Muhr, C., & Haslinger, D. (2014). A roadmap to continuous biometric authentication on mobile devices. In *Wireless Communications and Mobile Computing Conference (IWCMC), 2014 International* (pp. 122–127).: IEEE. (cited in 109).

[Ryu et al., 2006]
Ryu, C., Kim, H., & Jain, A. K. (2006). Template adaptation based fingerprint verification. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4 (pp. 582–585).: IEEE. (cited in 53 and 80).

[Rzouga Haddada, 2017]
Rzouga Haddada, L. (2017). *Contribution au renforcement du tatouage : Application à la sécurite des données biométriques et l authentification des individus*. PhD thesis, Ecole Nationale d Ingénieurs de Monastir. (cited in 25).

[Saidani et al., 2015]
Saidani, A., Kacem Echi, A., & Belaid, A. (2015). Arabic/latin and machine-printed/handwritten word discrimination using hog-based shape descriptor. *ELCVIA: electronic letters on computer vision and image analysis*, (pp. 0001–23). (cited in 25).

[Sang et al., 2004]
Sang, Y., Shen, H., & Fan, P. (2004). Novel impostors detection in keystroke dynamics by support vector machine. In *Parallel and distributed computing: applications and technologies* (pp. 666–669). Springer. (cited in 36).

[Scheidat et al., 2007]
Scheidat, T., Makrushin, A., & Vielhauer, C. (2007). Automatic template update strategies for biometrics. *Otto-von-Guericke University of Magdeburg, Magdeburg, Germany*. (cited in 59, 60, and 62).

[Schölkopf et al., 2001]
Schölkopf, B., Platt, J. C., Shawe-Taylor, J. C., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7), 1443–1471. (cited in 44).

[Scott Goldfine, 2015]
Scott Goldfine, Rodney Bosch, A. W. (2015). Security sales and integration. *EH Publishing*. (cited in 21).

[Seddik, 2017]
Seddik, B. (2017). *Multimodal Recognition of human-action streams: Application to sign language recognition*. PhD thesis, Sfax University. (cited in 26).

[Seddik et al., 2004a]
Seddik, H., Rahmouni, A., & Sayadi, M. (2004a). Text independent speaker recognition using the mel frequency cepstral coefficients and a neural network classifier. In *First International Symposium on Control, Communications and Signal Processing, 2004.* (pp. 631–634). (cited in 25).

[Seddik et al., 2004b]
Seddik, H., Rahmouni, A. B. S., & Sayadi, M. (2004b). Text independent speaker recognition based on the attack state formants and neural network classification. In *2004 IEEE International Conference on Industrial Technology, 2004. IEEE ICIT '04.*, volume 3 (pp. 1649–1653 Vol. 3). (cited in 25).

[Seeger & Bours, 2011]
Seeger, M. M. & Bours, P. (2011). How to comprehensively describe a biometric update mechanisms for keystroke dynamics. In *Security and Communication Networks (IWSCN), 2011 Third International Workshop on* (pp. 59–65).: IEEE. (cited in 45).

[Serwadda et al., 2013]
Serwadda, A., Balagani, K., Wang, Z., Koch, P., Govindarajan, S., Pokala, R., Goodkind, A., Brizan, D.-G., Rosenberg, A., & Phoha, V. V. (2013). Scan-based evaluation of continuous keystroke authentication systems. *IT Professional*, 15(4), 20–23. (cited in 49, 50, and 51).

[Shay et al., 2010]
Shay, R., Komanduri, S., Kelley, P. G., Leon, P. G., Mazurek, M. L., Bauer, L., Christin, N., & Cranor, L. F. (2010). Encountering stronger password requirements: user attitudes and behaviors. In *Proceedings of the Sixth Symposium on Usable Privacy and Security* (pp.2̃).: ACM. (cited in 15).

[Shih et al., 2018]
Shih, H.-Y., Guo, S., Chen, R.-C., & Peng, C.-Y. (2018). Enhanced passcode recognition based on press force and time interval. In *Asian Conference on Intelligent Information and Database Systems* (pp. 315–323).: Springer. (cited in 12).

[Smith et al., 2015]
Smith, D. F., Wiliem, A., & Lovell, B. C. (2015). Face recognition on consumer devices: Reflections on replay attacks. *IEEE Transactions on Information Forensics and Security*, 10(4), 736–745. (cited in 76).

[Stobert & Biddle, 2014]
Stobert, E. & Biddle, R. (2014). The password life cycle: user behaviour in managing passwords. In *Proc. SOUPS*. (cited in 14).

[Sukthankar & Stockton, 2001]
Sukthankar, R. & Stockton, R. (2001). Argus: the digital doorman. *IEEE Intelligent Systems*, 16(2), 14–19. (cited in 48, 50, 51, and 53).

[Taylor & Stone, 2009]
Taylor, M. E. & Stone, P. (2009). Transfer learning for reinforcement learning domains: A survey. *Journal of Machine Learning Research*, 10, 1633–1685. (cited in 60).

[Tsimperidis et al., 2018]
Tsimperidis, I., Yoo, P. D., Taha, K., Mylonas, A., & Katos, V. (2018). R^2bn: An adaptive model for keystroke-dynamics-based educational level classification. *IEEE Transactions on Cybernetics*, (pp. 1–11). (cited in 17, 27, and 41).

[Uludag et al., 2004]
Uludag, U., Ross, A., & Jain, A. (2004). Biometric template selection and update: a case study in fingerprints. *Pattern Recognition*, 37(7), 1533–1542. (cited in 52, 53, 56, 58, 59, 66, and 67).

[Ur et al., 2012]
Ur, B., Kelley, P. G., Komanduri, S., Lee, J., Maass, M., Mazurek, M. L., Passaro, T., Shay, R., Vidas, T., Bauer, L., et al. (2012). How does your password measure up? the effect of strength meters on password creation. In *USENIX Security Symposium* (pp. 65–80). (cited in 15).

[Vandana, 2007]
Vandana, K. (2007). Enhancing weak biometric authentication by adaptation and improved user-discrimination. Master's thesis, International Institute of Information Technology Hyderabad, INDIA. (cited in 48).

[Vibert, 2017]
Vibert, B. (2017). *Contributions to the evaluation of embedded biometric systems*. Theses, Normandie Université. (cited in 25).

[Vural et al., 2014]
Vural, E., Huang, J., Hou, D., & Schuckers, S. (2014). Shared research dataset to support development of keystroke authentication. In *Biometrics (IJCB), 2014 IEEE International Joint Conference on* (pp. 1–8).: IEEE. (cited in 28).

[Wang et al., 2012]
Wang, Z., Serwadda, A., Balagani, K. S., & Phoha, V. V. (2012). Transforming animals in a cyber-behavioral biometric menagerie with frog-boiling attacks. In

Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on (pp. 289–296).: IEEE. (cited in 48 and 51).

[Wilson, 1972]
Wilson, D. L. (1972). Asymptotic properties of nearest neighbor rules using edited data. *Systems, Man and Cybernetics, IEEE Transactions on*, 2(3), 408–421. (cited in 67).

[Xi et al., 2011]
Xi, K., Tang, Y., & Hu, J. (2011). Correlation keystroke verification scheme for user access control in cloud computing environment. *The Computer Journal*, (pp. bxr064). (cited in 28).

[Yager & Dunstone, 2007]
Yager, N. & Dunstone, T. (2007). Worms, chameleons, phantoms and doves: New additions to the biometric menagerie. In *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on* (pp. 1–6).: IEEE. (cited in 110).

[Yager & Dunstone, 2010]
Yager, N. & Dunstone, T. (2010). The biometric menagerie. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(2), 220–230. (cited in 111).

[Yu & Cho, 2003]
Yu, E. & Cho, S. (2003). Novelty detection approach for keystroke dynamics identity verification. In *International Conference on Intelligent Data Engineering and Automated Learning* (pp. 1016–1023).: Springer. (cited in 30, 35, and 36).

[Yu & Cho, 2004]
Yu, E. & Cho, S. (2004). Keystroke dynamics identity verification???its problems and practical solutions. *Computers & Security*, 23(5), 428 – 440. (cited in 36 and 75).

[Zhang et al., 2017]
Zhang, Q., Zhou, D., & Zeng, X. (2017). Heartid: a multiresolution convolutional neural network for ecg-based biometric human identification in smart health applications. *IEEE Access*, 5, 11805–11816. (cited in 26).

[Zhao et al., 2011]
Zhao, X., Evans, N., & Dugelay, J.-L. (2011). A co-training approach to automatic face recognition. In *Signal Processing Conference, 2011 19th European* (pp. 1979–1983).: IEEE. (cited in 64).

[Zhu et al., 2003]
Zhu, X., Ghahramani, Z., Lafferty, J., et al. (2003). Semi-supervised learning using gaussian fields and harmonic functions. In *ICML*, volume 3 (pp. 912–919). (cited in 57).

[Žliobaitė et al., 2015]
Žliobaitė, I., Bifet, A., Read, J., Pfahringer, B., & Holmes, G. (2015). Evaluation methods and decision theory for classification of streaming data with temporal dependence. *Machine Learning*, 98(3), 455–482. (cited in 52).