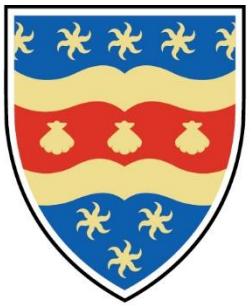
2019

Securing Cloud Storage by Transparent Biometric Cryptography

Abed, Leith



UNIVERSITY OF
PLYMOUTH

Securing Cloud Storage by Transparent Biometric Cryptography

by

Leith Hamid Abed

A thesis submitted to the University of Plymouth
in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing, Electronics and Mathematics

June 2019

Abstract

Securing Cloud Storage by Transparent Biometric Cryptography

Leith Hamid Abed

With the capability of storing huge volumes of data over the Internet, cloud storage has become a popular and desirable service for individuals and enterprises. The security issues, nevertheless, have been the intense debate within the cloud community. Significant attacks can be taken place, the most common being guessing the (poor) passwords. Given weaknesses with verification credentials, malicious attacks have happened across a variety of well-known storage services (i.e. Dropbox and Google Drive) – resulting in loss the privacy and confidentiality of files. Whilst today's use of third-party cryptographic applications can independently encrypt data, it arguably places a significant burden upon the user in terms of manually ciphering/deciphering each file and administering numerous keys in addition to the login password.

The field of biometric cryptography applies biometric modalities within cryptography to produce robust bio-crypto keys without having to remember them. There are, nonetheless, still specific flaws associated with the security of the established bio-crypto key and its usability. Users currently should present their biometric modalities intrusively each time a file needs to be encrypted/decrypted – thus leading to cumbersomeness and inconvenience while throughout usage. Transparent biometrics seeks to eliminate the explicit interaction for verification and thereby remove the user inconvenience. However, the application of transparent biometric within bio-cryptography can increase the variability of the biometric sample leading to further challenges on reproducing the bio-crypto key.

An innovative bio-cryptographic approach is developed to non-intrusively encrypt/decrypt data by a bio-crypto key established from transparent biometrics on the fly without storing it somewhere using a backpropagation neural network. This approach seeks to handle the shortcomings of the password login, and concurrently removes the usability issues of the third-party cryptographic applications – thus enabling a more secure and usable user-oriented level of encryption to reinforce the security controls within cloud-based storage. The challenge represents the ability of

the innovative bio-cryptographic approach to generate a reproducible bio-crypto key by selective transparent biometric modalities including fingerprint, face and keystrokes which are inherently noisier than their traditional counterparts. Accordingly, sets of experiments using functional and practical datasets reflecting a transparent and unconstrained sample collection are conducted to determine the reliability of creating a non-intrusive and repeatable bio-crypto key of a 256-bit length. With numerous samples being acquired in a non-intrusive fashion, the system would be spontaneously able to capture 6 samples within minute window of time. There is a possibility then to trade-off the false rejection against the false acceptance to tackle the high error, as long as the correct key can be generated via at least one successful sample. As such, the experiments demonstrate that a correct key can be generated to the genuine user once a minute and the average FAR was 0.9%, 0.06%, and 0.06% for fingerprint, face, and keystrokes respectively.

For further reinforcing the effectiveness of the key generation approach, other sets of experiments are also implemented to determine what impact the multibiometric approach would have upon the performance at the feature phase versus the matching phase. Holistically, the multibiometric key generation approach demonstrates the superiority in generating the bio-crypto key of a 256-bit in comparison with the single biometric approach. In particular, the feature-level fusion outperforms the matching-level fusion at producing the valid correct key with limited illegitimacy attempts in compromising it – 0.02% FAR rate overall. Accordingly, the thesis proposes an innovative bio-cryptosystem architecture by which cloud-independent encryption is provided to protect the users' personal data in a more reliable and usable fashion using non-intrusive multimodal biometrics.

Table of Contents

List of Figures

List of Tables

Acknowledgment

Most of all, I would like to thank Allah Almighty for giving me the health, knowledge, and capability to undertake the PhD study and to persevere in completing it satisfactorily. I do thank Him so much for His uncountable favour and without His help, this work would not have been possible.

I would like to express my grateful thanks to my Director of Studies Professor **Nathan Clarke** for his agile guidance. Special thanks also have to go to my second supervisor Dr **Bogdan Ghita** who has been supportive throughout my study.

I owe a debt of gratitude to my beloved parents my father (**Hamid**) and my mother (**Bahiya**) for their endless encouragement and support. Any success that might be resulted, optimistically, should help me making them proud and happy. I should not forget to thank my dear siblings who have been supportive without any hesitation, many thanks to them all.

My countless love, and appreciation must go to my wife (**Nada**), my son (**Wisam**), and my little daughter (**Laian**) who have been very patient and understanding throughout this journey, spending days, nights, and sometimes even holidays without me. I hope the potential success of this study will compensate some of what they have missed.

I am also very thankful to my best friends **Abdulrahman Alruban**, **Hussam Mohammed** and **Yaseen Alheety** for their assistance and support during the PhD endeavour.

Finally, I would like to express my sincere thanks to the Republic of Iraq and in particular the Higher Committee for Education Development for granting me a scholarship and sponsoring my PhD study.

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Doctoral College Quality Sub-Committee.

Work submitted for this research degree at the University of Plymouth has not formed part of any other degree either at the University of Plymouth or at another establishment.

This study was financed with the aid of a scholarship from the Higher Committee for Education Development (HCED) in the Republic of Iraq.

Relevant seminars and conferences were attended at which work was often presented and several papers are published:

- Abed L., Clarke N., Ghita B. (2017) Enabling Secure Cloud Storage through Transparent Authentication. Proceedings of the 8th Annual International Conference on ICT: Big Data, Cloud and Security (ICT-BDCS 2017), Singapore, 21st- 22nd August, pp 14-23, ISSN: 2251-2136.
  DOI: https://10.5176/2251-2136_ICT-BDCS17.07

- Abed L., Clarke N., Ghita B., Alruban A. (2019) Securing Cloud Storage by Transparent Biometric Cryptography. In: Lanet JL., Toma C. (eds) Innovative Security Solutions for Information Technology and Communications. SECITC 2018. Lecture Notes in Computer Science, vol 11359. Springer, Cham.
  DOI: https://doi.org/10.1007/978-3-030-12942-2_9

Word count of main body of thesis: 64,354 words

Signed .....

Date .....

Chapter One: Introduction

1.1 Overview

Cloud computing is an evolutionary paradigm in the scope of Internet-based computing providing services ranging from end-users applications, developers software platforms, to computing resources (Behl and Behl, 2012, Parekh and Sridaran, 2013). Amongst cloud computing services, cloud storage affords individuals and enterprises a free level of storage capacity for storing their own data on remote datacenters, aimed at abstracting away the complexity of hardware management and maintenance (Drago et al., 2012). In return for the immediate service provision, cloud storage providers charge the beneficiaries a very reasonable price per significant storage space (Ju et al., 2011). Customers also have ability to directly upload, download, update, remove, and share files via accessing their data from any-where at all times (Columbus, 2016). With the rapid increase in the amount of digital information, cloud storage has been a predominant service for storing data over the Internet (Phillipson, 2016). Therefore, this storage paradigm has become a very important topic in both academic and industrial communities (Behl and Behl, 2012). Microsoft OneDrive, Google Drive, and Dropbox are examples of the most popular and widespread cloud storage providers (Griffith, 2014). The number of Google Drive, and Dropbox subscribers world-wide has increased exponentially as illustrated in Figure 1.1 (Gannes, 2013, Sullivan, 2015, Columbus, 2016, Gildred, 2018).

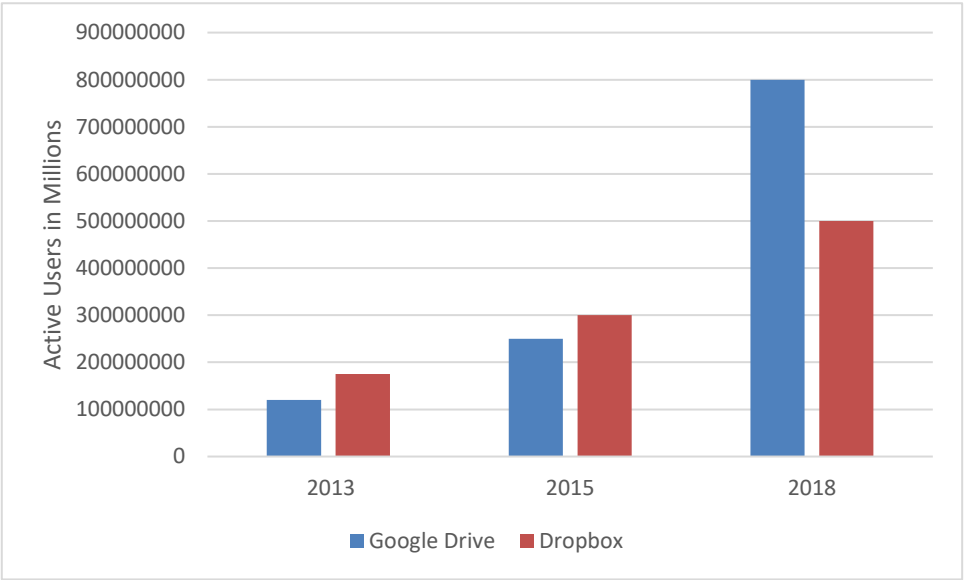


Figure 1. 1 The Growth of Cloud Storage Subscribers (adopted from Gannes, 2013, Sullivan, 2015, Columbus, 2016, Gildred, 2018)

According to the participatory-based study by Cloud Industry Forum (CIF), 88% of organizations in the UK utilized one of the cloud computing services, and the highest used cloud service was cloud storage (CIF, 2019). In the United States, business spent more than $13 billion on different cloud services in 2014 (McCue, 2014), and the majority of cloud consumers are expected to increase spending on storage services (Mellor, 2016). This shows that there is a very good sized market for the cloud computing services (Galibus et al., 2016) – and specifically for cloud storage services (Butler, 2013).

The cloud storage paradigm can be regarded as less secure than local storage, where the latter benefits from logical and physical security countermeasures before being able to access data. The local storage information can be stored on a hard drive in a computer on an office within an entire building. Thus, they are protected by both physical and logical security controls, such as secure door systems and firewalls, aimed at hindering the malicious attempts to hack the stored information.

In this case, an adversary either has to attack the entire building, or logically fudge around the firewall to violate stored data. As a result, there will be more challenging threat vectors to breach the stored information. However, within cloud storage-based systems, such as Dropbox and Google Drive, potential attacks can simply take place through the web portals by which users log in to get access to their accounts. Of course, cloud systems have stronger security measures including fingerprint locks, and armed guards for protecting and monitoring the data storage centres preventing attackers from violating them (Dobran, 2019). Nevertheless, the access into the cloud storage data is still granted using password verification and guessing the potential weak password can leak the user account from single portal and this is not the case within the PC login. Millions of user accounts are also accessible via the same portal - providing a single point of attack. As such, the attacks on local storage technologies can be considered more difficult than cloud-based storage.

There is no doubt that cloud storage provides subscribers highly advantageous attributes including scalability, flexibility, accessibility, usability and data backup (NT, 2014). However, security issues have been the intense debate within the scientific communities (Parekh and Sridaran, 2013, Behl and Behl, 2012). Overall, cloud providers employ a number of security countermeasures, aimed at achieving robust security. Data in transit is secured by standard Internet security protocols, such as Transport Layer Security (TLS). Cryptographic techniques also encrypt data at rest to protect them from external attacks (Galibus et al., 2016). Accessing the stored data is granted by successful password verification (Azam and Johnsson, 2011). If an attacker can obtain a verification credential, neither the security controls in transit nor at rest can combat that attack. This results in breaching data confidentiality –

since the malicious access is considered to be genuine with successful verification. Even if two-factor authentication is used to mitigate this vulnerability, there are still simple passphrases can be hacked (Toorani and Beheshti, 2008). In addition, this approach is not universally adopted and providing further burden upon the user to have a second login using another piece of credential. Passwords techniques have been the topic of intense debate in academia, and are arguably manifested to be often poor (Uludag et al., 2004, Rathgeb and Uhl, 2011). Subscribers commonly access cloud services by simple passwords. As such, a number of recent attacks have had significant impact. For instance, approximately 7 million Dropbox accounts were leaked illegally (Kovach, 2014), and around 5 million Google Drive accounts were hacked in 2014 (Vonnie, 2014). The selection of poor passwords arises when users are struggling with recalling, managing, and using complex passwords. With a view to tackling the current security issues, many clients have sought to provide additional security through using third-party encryption tools to manually encrypt or decrypt data prior to putting it into the cloud (Bischoff, 2018). Illegitimate access to the cloud service will result in violating encrypted files which still require a secondary key – ideally for each and every file to breach them. However, these tools still bring usability issues in terms of having to manage a key for each file in addition to the login password. Subscribers also need to cipher/decipher each file manually – further exacerbating the usability issue.

From the above presentation, it is clear that the current security approaches are not providing adequate security without introducing significant usability issues. Biometric cryptography produces reliable and usable bio-crypto keys from biometrics modalities (Kanade et al., 2009b) – where there is no need to remember complex

passwords/keys. In essence, the bio-crypto key can be established by constructing and storing public data from biometric on registration, and this data will be used on verification to reproduce the same key for cryptographic goals, such as encryption and authentication (Rathgeb and Uhl, 2011). Bio-cryptography accordingly ensures that sensitive information (i.e. biometric information, secret keys) will not be stored somewhere within bio-cryptosystem. Thus, bio-cryptographic approaches overcome significant security attacks that happen upon the storage component within cryptographic and biometric systems (Uludag et al., 2004). In addition, the traditional password authentication will no longer need to protect the secret keys. Given the characteristic of storing public data only, bio-cryptography can also revoke the security credentials (biometric templates, secret keys and passwords) in case of compromise (Rathgeb and Busch, 2012). As such, the application of bio-crypto keys within cloud storage can offer the capacity to manage the above-mentioned security and privacy issues without incorporating any auxiliary applications. That is, a file will be only decrypted by the bio-crypto key that is established using the biometric features of a genuine user as these features have a high level of uniqueness to a distinct person. The attacker who is attempting to hack the file will have no capacity to produce the same key as the presented biometric features are different. Even if an attacker had the ability of guessing a weak password, he cannot hack this file because it is encrypted by a key established from the biometric features of the cloud user. However, there are still some impairments related to the usability of the bio-crypto key and its security in practice. Users currently need to present their biometric modalities intrusively each time a file needs to be encrypted or decrypted; thus, it still presents a usability issue. On the other hand, achieving a repeatable bio-crypto key

in a secure manner with the absence of storing biometric information can be considered very challenging as biometric features are inconsistent over time.

Using the research area of transparent and continuous biometric verification, in which biometrical signals are collected in a non-intrusive fashion, offers the opportunity to remove many of the usability issues associated with traditional biometric cryptosystems – potentially enabling more usable and secure cryptography. However, an effective bio-crypto approach that can successfully compass transparent biometric in a secure manner needs to be identified.

1.2 Research Goal and Objectives

The aim of this research is to develop an innovative bio-cryptographic approach using transparent biometrics in order to reinforce the lack of security controls within cloud-based storage. A transparent encryption framework built upon this approach would maximize the level of protection and convenience – the user no longer needs to recall, or present complex credentials, and the encryption is seamlessly undertaken for the authorized identity.

With a view to accomplishing the research goal, subordinate research objectives are established in order to:

- Review the current state-of-the-art in the research area of biometric cryptography in terms of existing approaches, strategic schemas, issues and available solutions.
- Conduct a number of investigations to explore whether a repeatable bio-crypto key can be established from a single transparent biometric throughout.

- Perform a set of experiments to investigate the effectiveness of biometric fusion approaches for producing a superior repeatable bio-crypto key from multiple non-intrusive biometric modalities on a timely basis.
- Design and develop an innovative transparent and multimodal bio-cryptosystem architecture for cloud storage technology capable of providing a secure, robust and frictionless user experience.

1.3 Thesis Organization

In addition to this chapter, which presents the research problem, the overall aim and objectives, and the structure of the research, the thesis contains a further six chapters outlined as follows:

Chapter 2 is titled "Biometric Systems". This chapter introduces the theoretical background of the research. This mainly includes the basics of biometric systems, the concepts of continuous and transparent biometric authentication, and the principles of biometric cryptography. The biometric system fundamentals are elaborated in terms of biometric requirements, system components, and performance measurement. A devoted section is also presented to discuss multibiometrics and in particular the levels of fusion in which biometric modalities can be consolidated. Accordingly, a conceptual view of transparent biometric verification is illustrated with a concentration upon contextualizing its approaches toward the core of this work to handle security and usability issues. The principles of biometric cryptography are ultimately considered and explained with the purpose of understanding to what extent such an approach can robustly tackle the security and privacy issues of cloud storage.

Chapter 3 presents a critical analysis of the current state-of-the-art comprehension of biometric encryption. The review of biometric cryptographic approaches is broken down into several thematic sections, and the research is presented in a chronological order.

Chapter 4 initially proposes a novel bio-cryptographic approach for enabling a more secure and usable cloud storage through transparent biometric modalities, given the lack of additional protection in place. Accordingly, a set of essential investigations are undertaken and carried out aiming to ultimately discover the potential contribution of the developed approach. The first series of experiments concentrates upon investigating how reliable the innovative bio-cryptographic approach in generating a bio-crypto key from transparent biometric modalities. Another set of experiments explores the potential of enhancing the performance of the bio-crypto key generation. The final experiments seek to investigate the capacity of generating different cryptographic key sizes through biometric features.

Chapter 5 is titled "**Investigation into Transparent Multibiometric Cryptography**". This chapter seeks to develop an advanced bio-cryptographic model using the principle of multibiometric fusion, aiming ultimately at investigating to what degree it can improve the performance of the bio-crypto key generation over the single biometric modalities. Accordingly, two fundamental investigations are developed and conducted in order to explore the potential outperformance via multibiometric. The former experiment determines the key generation effectiveness at the feature level, while the latter explores the performance at the matching level.

Chapter 6 presents the architectural framework of the innovative multibiometric cryptosystem. The essential system architecture requirements are primarily

identified depending upon the obtained knowledge and the experimental outcomes from the previous chapters. Then, a comprehensive clarification of the system components, and mechanisms is presented - with a concentration upon tackling the security and usability issues in order to ensure a convenient and reliable experience for cloud storage subscribers. A number of operational considerations are also addressed and conceptually explicated with a view to reinforcing the system operation in practice.

Chapter 7 presents the fundamental conclusions arising from the research; the main contributions, achievements and limitations. It also poses a discussion on potential areas for future work.

Chapter Two: Biometric Systems

2.1 Introduction

For many years, human traits (biometrics) have been employed to identify people; for example, individuals can be recognized via their fingerprints, irises, and voices (Jain et al., 2007). In seeking a secure solution, biometrics modalities have been applied to achieve sophisticated authentication and identification systems. Biometrics approaches can provide a reliable protection for environments that require high level of security as biometric features are very unique to an individual (Clarke, 2011). Biometric identifiers can also prevent the person from having to remember/recall difficult credentials or carry and protect tokens (Jain et al., 2007). Biometrics approaches, therefore, have been adopted in a variety of applications, including border agencies and military organizations (Soutar et al., 1999). After five decades of research in biometric, its techniques have been widely developed in the last years – where they are built in various everyday technologies, such as mice, keyboards, laptops, smartphones, and ATMs. As a result, the biometrics store is expected to grow over 304% between 2016 and 2023 to exceed $34.60-billion (MarketsandMarkets, 2016).

Traditional biometric authentication approaches verify the user at first of the session only not frequently (i.e. point-of-entry verification) which resulting in particular shortcomings. That is, when the genuine person has gone away from his active device after a successful verification, critical vulnerabilities can arise (Clarke and Furnell, 2005). Biometric techniques , on the other hand, can be exploited farther to authenticate the user frequently via capturing the biometrics traits in a non-intrusive

fashion without any inconveniences – thus defeating the initial verification only at the beginning of the session and overcoming the flaws of traditional verification at point-of-entry only (Clarke, 2011). In particular, the transparent approaches, such as face and voice afford users a more usable and secure way in practice, where there is no explicit interaction for the continuous verification against imposters. In addition, as the biometric samples are collected in a spontaneous manner, an adversary will have difficulties in spoofing the collective biometric signals (Clarke, 2011).

From a cryptographic perspective, the existing encryption approaches lack the secure management of the secret keys. Fundamentally, the cryptosystems can be classified into two systems: symmetric key and asymmetric key systems (Soutar et al., 1999). Symmetric key systems use a symmetrical key for encryption and decryption, and it must be securely stored somewhere. Asymmetric key systems, however, use public and private keys. The public one is utilised for ciphering the secret information, and it is distributed amongst the communicative parties for verification aims. On the other hand, the private key is utilized to decipher the information, so it has to be stored in a secure place as well. On the whole, there are two issues with the above cryptographic systems. First, adversaries could attack the private key transmission from one party to another. Second, potential attacks may take place on the stored private key. Consequently, a mechanism is required to cope with these security issues. Whilst Date in transit can be protected via standard Internet security protocols (e.g. Transport Layer Security (TLS)) (Galibus et al., 2016), the access to the stored secret keys is permitted via traditional password authentication approaches (Chang, 2012). Overall, passwords are weakly selected, and often derived from personal information, and this will make the system

vulnerable to several attacks, such as password guessing attacks (Kanade et al., 2008). In addition to this, cryptography does not ensure that an individual, who provides the password, is a legitimated user (Soutar et al., 1999). Consequently, Bodo (1994) innovated the idea of establishing robust encryption based on biometrics due to its capability of identifying human beings in a reliable way. Thus, a number of efforts in combining biometrics and cryptography have resulted in developing the field of biometric cryptography. In particular, bio-cryptographic techniques seek to establish secret keys from biometrics in a secure management in which nor keys neither biometrics would be stored somewhere. As such, the biometric keys can be revoked in case of compromise. Contrarily, password-based techniques need always to store the passwords at some location and accordingly, they cannot be cancelled if they have been hacked (Cavoukian and Stoianov, 2007). In addition, whilst bio-cryptography offers strong bio-crypto keys without forcing the users to remember them, weak passwords can be selected when users are struggling with recalling, managing, and using complex passwords. Table 2.1 summaries the security criteria between bio-cryptographic techniques and password-based techniques.

Table 2. 1 The Security Criteria of Bio-Cryptographic Techniques versus Password-Based Techniques (adapted from Cavoukian and Stoianov, 2007)

Bio-cryptographic Techniques	Password-Based Techniques
Neither key nor biometric should be stored	A password should be always stored
A hacked key can be revoked	A hacked password cannot be revoked
Key is strong	A password can be strong or weak

From the above presentation, it is clear that the use of biometric encryption seems to be more robust than passwords for protecting secret keys and/or sensitive information. Employing transparent biometric could also present the capacity of

eradicating particular vulnerabilities and inconveniences relevant with traditional bio-cryptosystems – probably providing more secure and usable bio-cryptographic framework for cloud-based storage. So as to introduce an insight into transparent bio-cryptography, this chapter states the biometric systems in terms of characteristics, performance metrics, advantages and disadvantages. Furthermore, the capacity of applying a number of biometric techniques in a transparent mode is discussed for tackling the usability issues within cloud-based storage. Biometric cryptography is also explained with regard to the overall concept, the approaches of bio-cryptosystems, system requirements and performance measurement.

2.2 Biometric Characteristics

The selection of biometric modalities for security purposes is dependant on a number of characteristics which are very important to be taken into account. Consequently, biometric modalities can be considered suitable for security applications, once they are met all of the requirements of universality, uniqueness, permanence, collectability, circumvention, performance and acceptability. An explanation for each characteristic is listed below (Jain et al., 2007):

- **Universality** means that the exploitable biometric identifier needs to be available over the complete population of users.
- **Uniqueness** refers to the level of distinctiveness of the biometric characteristic. That is, any two persons should be completely different with regard to their biometric traits for successful verification.
- **Permanence** represents the capability to generate a stable biometric template over time. For instance, the iris is one of the most consistent

biometrics over very long periods of times, but gait behavioural recognition can be inconsistent since the person could be in hurry or tired.

- **Collectability** indicates the flexibility to collect the chosen modality for a biometric system. Specific biometrics approaches can be regarded very intrusive where they require special devices and/or explicit user interaction, such as a retina scan. On the other hand, some biometric approaches can be captured easily with normal daily devices and interactions, such as gathering voice samples when having a phone call.
- **Circumvention** implies to how difficult it is to attack the biometric system by imposters.
- **Performance** refers to the scalability of the biometric features to meet the specific achievement of accuracy and reliability. The performance of the biometric system is increased whenever the biometric features are constant over time (Atah, 2011).
- **Acceptability** indicates to what extent people that are willing to accept the biometric system.

2.3 Components of a Biometric System

A conventional biometric system mainly comprises of five components which are described as below (Jain et al., 2007, Clarke, 2011):

1. **Acquisition**

   The biometric patterns are captured from an individual by a viable acquisition device. Some biometric technologies can use existing equipment such as

face recognition via webcam whilst others need sophisticated scanners (e.g. eye retina recognition).

2. Feature Extraction

   Through this component, the biometric features are extracted from the captured samples using particular signal processing algorithms to generate a source biometric template.

3. Storage

   The source biometric template is stored on a database or a smartcard in order to be used in the matching process.

4. Matching or Classification

   With the purpose of achieving secure access for a genuine user, biometric features are extracted from live biometric samples to compare against the stored template by using a matching algorithm. Consequently, their degree of similarity is represented by a match score.

5. Decision

   The degree of required similarity which results in a yes or no response is usually predefined in the system by a threshold. Depending on the response of the system the access will be granted or denied. Figure 2.1 shows the components of the common biometric system:

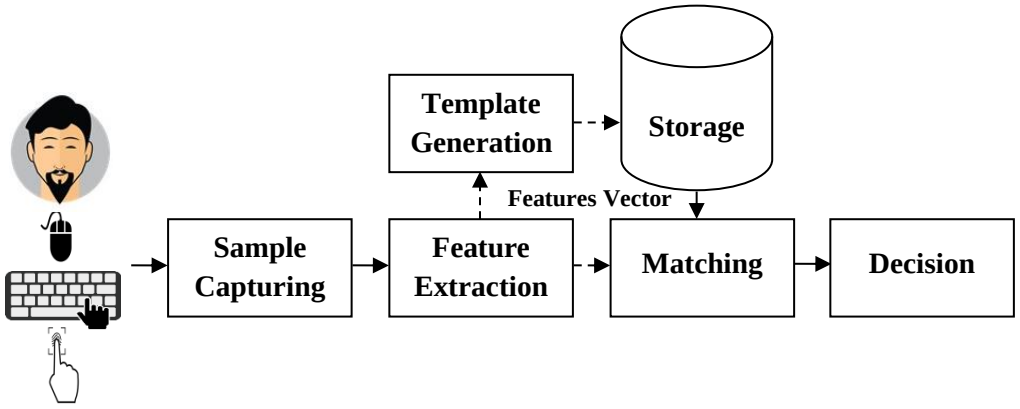


Figure 2. 1 The Components of the Biometric System (modified from Clarke 2011)

On the whole, a biometric authentication is achieved through the enrolment and the verification stages. At the time of enrolment, biometric samples of the legitimate are acquired, and then a biometric template is generated by applying a number of pre-processing and feature extraction algorithms. This biometric template will be stored for subsequent use – where it would be compared against future samples. The quality of biometric samples should be ameliorated to ensure that they are appropriate for successful verification. On authentication, the biometric system verifies the current identity of user with his/her stored identity. The access to the system will be granted to the user if the matching process results in a similarity score that is sufficiently high. The decision component determines what the threshold for acceptance is (Jain et al., 2004).

Biometric approaches fall into two groups: physiological and behavioural biometrics. Physiological approaches rely on the human being body, such as the shape of a face, or eye. Behavioural approaches, however, depend on the behaviour of users when they are doing specific tasks, such as walking or speaking (Ashbourn, 2004). Overall, physiological biometric techniques tend to be the more reliable and mature technology, and they are widespread used due to the tendency of their biometric

features to be invariant over time (Le and Jain, 2009). According to the participatory-based research by Biometrics Institute Industry (BII), the form under which the biometric technologies are classed on the basis of the participant preference starts with fingerprint then face followed by iris (physiological approaches). However, speaker recognition (behavioural approach) comes at the end of the study list (BiometricsInstituteLimited, 2013). Behavioural biometric approaches, however, are typically more convenient than physiological biometrics in terms of collection (Clarke, 2011). That is, the authentic user may not have to react with behavioural biometric system through the sample acquisition phase. The explicit interaction for biometric verification can normally take a while each time – leading to a tedious process. For instance, the voice samples can be captured in a passive manner when the user has a phone call.

2.4 Performance Measurement of Biometric System

As mentioned previously, a biometric system recognises the genuine person from others through the matching or comparison between the target biometric template of the current sample and the stored enrolment biometric template. During this time, several factors can impact the performance of biometric system. For instance, environmental noises can prevent legitimate users to access the system, and conversely allow impostors to get access when they should not.

There are two error rates that can evaluate the accuracy of biometric systems. These error rates are the False Rejection Rate (FRR), and the False Acceptance Rate (FAR). FRR measures the rate of biometric system error in rejecting genuine users; however, FAR measures the rate of biometric system error in accepting forgers

(Ross et al., 2006, Jain et al., 2007, Clarke, 2011). Figure 2.2 shows the metrics of biometric system performance in terms of FRR and FAR as below:

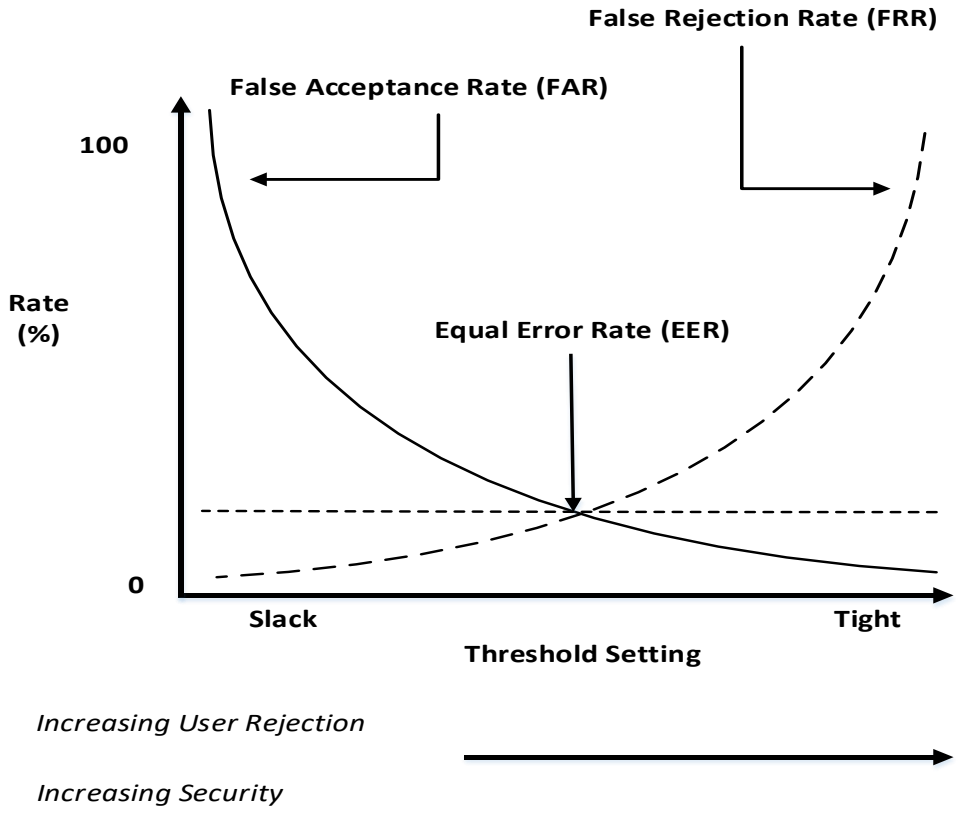


Figure 2. 2 The Metrics Performance in terms of FRR and FAR (Clarke, 2011)

The acceptable values of FRR and FAR are managed by the predefined threshold. Depending on this threshold, the requirements of the system security and usability can be determined. As shown in Figure 2.2, when the threshold is defined tight – requiring a significant level of matching, the access of authentic users to their accounts could be ignored (FRR) (Clarke and Furnell, 2005). In this case, the security level can be underpinned in terms of rejecting illegitimate access (i.e. the FAR will be decreased). However, this would not necessarily improve the overall biometric security. This can also result in user inconveniences due to the repetitive unsuccessful verifications - such systems may not get attracting to be adopted. On the other hand, when the threshold is defined slack – requiring a low level of matching,

the possibility of forger's access to the biometric system will be raised (FAR). Although this would achieve a very convenient verification to valid users by reducing the likelihood of being refused (FRR), it will influence the security aspects negatively. Consequently, the aspects of security and usability should be cautiously balanced. As depicted in Figure 2.2, the accuracy of the biometric system can be also evaluated by another metric called Equal Error Rate (EER) in which the FRR and FAR are intersected. Whenever the EER value decreases, the performance of a biometric system increases and vice versa (Clarke, 2011).

2.5 Multibiometrics

Generally, all biometric approaches can be operated in a single or multimodal biometrics. There is no doubt that the system of multibiometric copes with certain shortcomings of single biometric systems. Single biometric systems have a number of weaknesses, such as low individuality, high forgery attempts, high error rate, and lack of universality. For instance, in face recognition, it is impacted by position, expressions and the amount of present illumination. Also, it has been evident for most contributors that around 2% of the population does not have a legible fingerprint. As a result, they cannot be registered into a fingerprint biometrics system (Congress, 2012). Thus, unimodal biometric can be inadequate for some cases and individuals.

The multibiometric system, however, can offer more secure and convenient aspects for particular populations and applications (Ross et al., 2006). For example, national border agencies require applications of high security that can be offered by multi-biometric system. In this context, the AND bitwise fusion of two or more subsequent verifications are applied to be more difficult to tamper with by forgers (Cimato et al.,

2009). Another instance is when users are incapable of providing a specific biometric sample either temporarily (e.g. a person with broken hand cannot have hand geometry samples) or permanently (e.g. a wheelchair person cannot have gait samples). In this situation, the OR bitwise fusion can increase the population coverage (Cimato et al., 2009). Consequently, multibiometric systems could offer a higher level of flexibility, convenience, and security over their single biometric counterparts (Ross et al., 2006). Whilst this could improve the system accuracy, robustness and reliability, considerations, such as processing load, cost, and vendor-services should be taken into account prior to deploying such a system.

On the whole, multibiometrics systems can be developed by utilising one of the following sources (Ross et al., 2006):

- Multimodal in which multiple biometric modalities are used, such as voice and face or iris and gait.
- Multi-instance where more than one subtype of the same biometrics is utilised, such as the right and left iris biometrics.
- Multi-sensor means that multiple sensors are exploited to acquire a single biometric of a person, such as using both optical and capacitive fingerprint sensors.
- Multiple samples under which a single sensor is used to capture more than one sample of the same biometric with taking into account of their potential variations, such as face poses.
- Multiple algorithms mean that more than one classification algorithm upon a single biometric is used to combine the resultant features (e.g. minutiae-based and texture-based fingerprint classifier algorithms).

- Hybrid by where a subset of the above-mentioned categories is exploited, aiming at improving the recognition accuracy. For example, three face recognition algorithms can be incorporated with two iris recognition algorithms.

The variety of information sources (multimodal, multi-instance, multi-sensor, multi-sample, multi-algorithmic, and hybrid approaches) for a multibiometric system aims to improve the verification decision. Therefore, the way of combining the biometric features, which is termed fusion, should be employed carefully to reinforce the decision process. Overall, the fusion method can be applied during particular levels in the biometric system. These levels that are sensor, feature, matching and/or decision level are illustrated as below (Ross et al., 2006, Sim et al., 2007):

- Sensor level fusion integrates the raw data of multiple biometric samples afore the feature extraction stage. The raw data can be captured by a single sensor or by multiple sensors (e.g. consolidating several iris images from one or several sensors).
- Feature level fusion consolidates multiple feature vectors that are extracted from the samples of one or more biometric modalities by using several feature extraction algorithms. The fusion of multiple feature vectors will be used in the matching phase (e.g. consolidating the feature vectors of the fingerprint and iris).
- Score level fusion in which the results of multiple biometrics matchers are combined to obtain a new stacked match score that will be exploited for the consecutive decision process.
- Decision level fusion occurs when each associated biometric system has presented its own decision to provide a final decision.

2.6 Continuous and Transparent Biometric Authentication

Most standard authentication approaches, such as biometric authentication establishes initial user verification at the beginning of the session only not frequently (i.e. point-of-entry authentication); accordingly, these approaches have particular shortcomings. That is, when point-of-entry authentication has been successfully achieved, and the genuine person has gone away from his active device for significant periods of time, critical issues can arise (Clarke and Furnell, 2005). Serious vulnerabilities under which an adversary can attack the device will take place after an initial genuine login only – where free and open abuse can be performed. The majority of the authentication approaches verify the valid user at the time of making the access control decision only not throughout as shown in Figure 2.3 (Clarke, 2011):

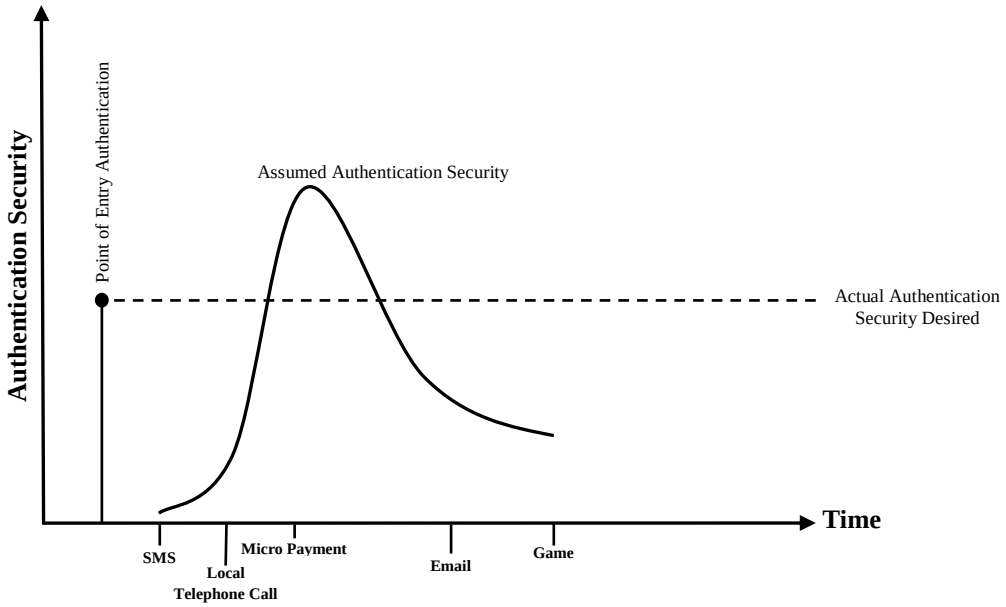


Figure 2. 3 A Model of Conventional Authentication (Clarke 2011)

As such, it is arguably advantageous to maximise the authentication of users beyond simple and standard authentication approaches. A possible direction of coping with the misuse issues is to apply sophisticated authentication techniques that can apply user reverification continuously and periodically without any inconveniences. Most of the applicable authentication techniques are built in an intrusive fashion. To say, there is an explicit interaction between the user and the system. For instance, the user should type the password to gain access to the system (Ceccarelli et al., 2015). Although biometric authentication is intrusively deployed, it can be developed in a more usable and secure manner. That is, biometrics can be collected in a spontaneous or non-intrusive way, aimed at verifying the legitimate users continuously. Therefore, transparent or non-intrusive biometric is an approach in which an active and continuous verification mechanism is provided over time by non-intrusively collecting the biometric samples thus eradicating the inconveniences on having to explicitly interact with the system. This authentication presents the opportunity of immigrating a yes/no response to a more appropriate and reliable

decision where a non-intrusive authentication process is more closely stood up with the access control decision. Moreover, transparent biometric verification takes into account that the entire authentication techniques are unequal, and they have different levels of effectiveness as illustrated in Figure 2.4 (Clarke, 2011):

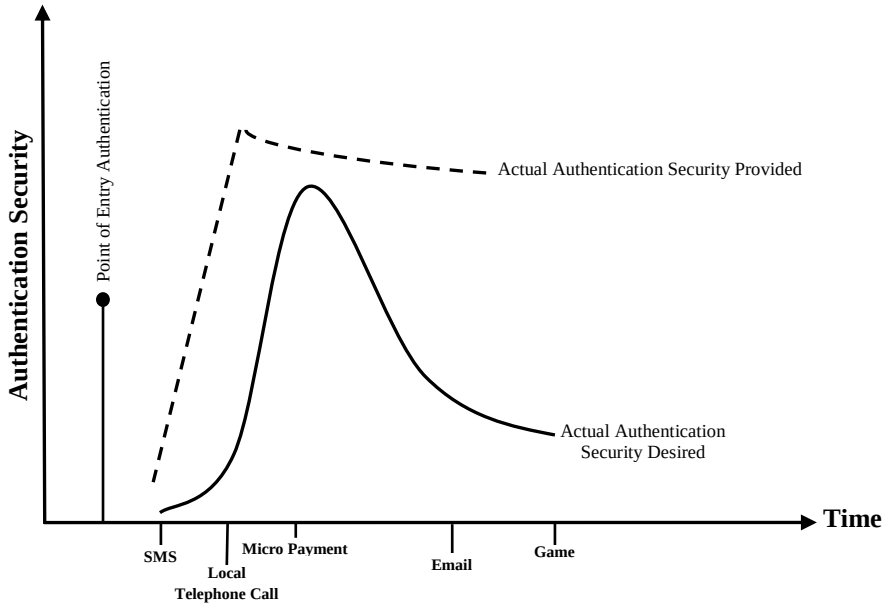


Figure 2. 4 A Model of Continuous Authentication Confidence (Clarke 2011)

However, the establishment of transparent authentication is still a challenging task. Although transparent authentication presents a continuous and flexible authentication over time, specific issues impede the verification process. These largely concentrate around the case of acquiring a person's biometric sample in a non-intrusive way. For instance, in the case of capturing the face samples, external and environmental factors, such as the distance of the camera, face orientation and lighting considerably impact the accuracy of facial detection (Clarke, 2011). Furthermore, not all biometric techniques can be adapted to operate in a transparent manner. This chapter will concentrate in the next subsection on discussing biometric

techniques that are considered transparent-enabled approaches and appropriate with the context of this research.

In spite of the abovementioned barriers and challenges of transparent authentication, its conception can offer potential benefit over traditional intrusive biometric in this research. Applying transparent biometrics within bio-cryptography can remove many of the usability issues of the existing cloud-based storage model and simultaneously introduce cloud-independent encryption layer. The cloud subscriber will no longer need to manually encrypt/decrypt each and every file prior to putting it into cloud storage through auxiliary encryption tools. In addition, the encryption key will be biometrically established on the fly without storing it somewhere to overcome potential attacks upon stored security credentials (biometrics, secret keys/passwords).

2.6.1 Transparent Biometric Approaches

A number of biometric modalities can be considered transparent-enabled approaches as they do not necessarily require an explicit interaction for sample collection. Transparent biometric techniques can be also classified into two categories: physiological (e.g. face, ear and fingerprint) and behavioural, such as keystrokes analysis, eye gaze or eye tracking and gait (Clarke, 2011). Of course, the features in both categories are established in non-intrusive manner without inconveniencing the user. A variety of physiological and behavioural approaches that could be transparently employed in the context of this research are debated in the following subsections. These subsections start with physiological biometrics of

fingerprint, face, eye geometry and ear, and then turn into behavioural biometrics of voice, keystrokes analysis and behavioural profiling.

2.6.1.1 Fingerprint Recognition

Fingerprint recognition is the oldest and popular biometric technologies that have been widely used in many computing and mobile applications for authentication purposes. The adoption of fingerprint technology become very prevalent around because profound studies have been experimentally proven that fingerprint recognition approach has high level of individuality to each finger (Jain et al., 2007). As a result, the applications of fingerprint biometric are exploited for both physical and logical access control. Overall, fingerprint recognition compares ridges, valleys and patterns of a person fingerprint via one of the three matching classification approaches: minutiae-based, correlation-based and ridge feature-based (Maltoni et al., 2009). It is evident that there are no two individuals (including identical twins) sharing the same fingerprints as the fingerprint patterns are part of an individual's phenotype, and do not seemingly rely on genetics (Danielyan, 2004, Ross et al., 2006).

Fingerprints have been used to recognise human beings for a long time. During 1926, law enforcement in American cities had started submitting fingerprint cards to the Federal Bureau of Investigation (FBI) in an attempt to build a database of fingerprints from known criminals. In the early 1960s, fingerprint technology became automated with the Automated Fingerprint Identification System (AFIS) (Soutar et al., 1999). Later, the FBI developed the AFIS in 1999 with respect to response time and capacity, and at the same time included the ten fingerprints to become the Integrated

Automated Fingerprint Identification System (IAFIS). Despite the evidences of pre-historic use of fingerprint technology for human recognition (Holder et al., 2011), the current applications of fingerprint technique earned wide prominence since the evolution of digital computer, and has continued to remain popular in forensics, commercial, civil and Government applications.

From the unconstrained authentication perspective, fingerprint has the potential to be applied in a transparent fashion. There are some efforts that can reflect promising indications towards achieving mature transparent fingerprint techniques. Owing to the rush in the number of touch based smart devices, there is an increasing necessity for developing a convenient authentication framework via the fingerprint uniqueness of person's identity. As such, fingerprint samples can be collected by using capacitive sensing techniques in touchable smart devices. In an interesting review, Koundinya et al. (2014) proposed an innovative integrated device using transparent electronics for both multi-touch interaction and fingerprint scan. Non-intrusive touch sensitive device and an input/output circuit which drives the capacitive sensor array for fingerprint sensing at higher resolutions and for touch interactions at low resolutions are discussed in an elaborate manner. The experimental results demonstrated that the proposed scheme can be recognised, and can present a highly efficient means for transparent user authentication. Furthermore, Apple Company has recently introduced the Touch ID fingerprint scanner on the iPhone 6 (Hattersley, 2016). Using this technology, a person should initially register his finger samples to the system. As such, whenever a person places his finger on the home button to unlock the mobile phone device, the hardware passively scans the finger, and the software establishes the genuine verification. Apple's implementation of the Touch ID can

arguably reflect transparent and continuous authentication (Hattersley, 2016). It can be argued that the novel insights of employing the fingerprint modality towards unconstrained sample collection will undoubtedly present mature transparent and continuous fingerprint verification for enhancing the security and usability aspects.

2.6.1.2 Face Recognition

Facial recognition has been widely utilised in the computer security and surveillance applications since it can provide vital discriminative features for recognising human beings. (Jain et al., 2007). The face modality is deemed a passive biometric since it does not necessarily require the reaction of the user to achieve recognition as well as the unobtrusive nature of its technology makes it an attractive choice for many security applications. For example, an automated face recognition system can use a video camera to capture face images from a distance, and detect, track and finally recognize people, such as terrorists or drug traffickers (Solayappan and Latifi, 2006).

The features of the face biometric can be the dimensions of the eyes, nose, mouth, ears, cheekbones, and the distance among most or all of them. The location and shape the face attributes (e.g. eyes, nose and lips) can be also used as facial features. These facial features are extracted by particular algorithms, and their effectiveness depend on several factors, such as the consistency of the extracted features over time, image resolution, surrounding lighting, and face distance and position from the camera (Clarke and Furnell, 2005). As a result, certain schemes have been presented in order to manage some of these factors in a transparent mode. Using a three-dimensional image modal could assist in mitigating the impacts of face orientation and illumination conditions. However, the need for 3D acquisition

device, such as camera or sensor would hinder its acceptance (Clarke, 2011) due to their tendency to be slower in response and more expensive (Jain et al., 2007). What is more, Clarke et al. (2008) transparently suggested an advanced composite paradigm in which a number of person's face samples in different sizes, orientations and illumination are jointly stored as a biometric template. Accordingly, when a sample is captured, it will be compared with the stored composite template to accomplish user authentication. For example, if the taken sample is facing down, then it will be matched with the stored template under which that orientation exists and similarly for other instances. On the other hand, the trade-off between security and user friendliness is an issue because the possibility of rejecting a legitimate user decreases while that of accepting a forger increases.

2.6.1.3 Eye Geometry

As one of the distinctive facial features, human eyes specifically play a vital role in face recognition and facial expressions analysis. The eye geometry is mostly more prominent and consistent feature than the other facial features over time (Peng et al., 2005). As such, the landmarks of eye patterns are considered an important discovery for quick, convenient, and reliable pattern recognition since they are quite unique. As a result, the eye recognition technology could be close to the other biometric technologies in terms of performance, such as iris, voice and fingerprint recognition. However, it seems that there have been no independent trials of the eye geometry recognition technology (Panda and Ranjan, 2007). On the whole, eye recognition technique depends on the eye detection approaches (Peng et al., 2005). Various papers, therefore, have been published in the scope of eye detection in order to present the capacity of developing eye recognition.

Eye detection approaches in the literature can be divided into two groups: active infrared (IR) approaches and passive approaches. The detection approaches which depend on active remote IR illumination are simple and effective, where they utilise an active IR light source to obtain the bright or dark pupil impacts. That is to say, these methods can be only applied to the IR illuminated eye pictures. Of course, these methods are not widely used as the eye samples of real-time applications are apparently not IR illuminated (Peng et al., 2005). On the other hand, the passive methods can be divided into three types: template based methods (Xie et al., 1994), appearance based methods (Huang and Wechsler, 1999) and feature based methods (Feng and Yuen, 1998). In template based methods, a generic eye model, which is based on the eye shape, is designed first. Accordingly, template matching is used to find the nearest picture for the test samples. Whilst these methods can accurately detect eyes geometry, they are taking considerable period of time to process the eye detection (Peng et al., 2005). The appearance based methods can detect eyes based on their photometric appearance. However, these methods usually need to collect a large amount of training data, representing the eyes of various volunteers, under different facial orientations, and under different illumination conditions. So as to accomplish the eye detection process, these training data are used to learn a classifier, such as a neural network or the support vector machine, and depending on the matching score, the decision is made (Huang and Wechsler, 1999). Feature based methods discover the eye features of edge and the intensity (i.e. the colour distributions of the sclera and the flesh of the eyes) to identify unique biometric features. Even though these methods are usually efficient, they lack accuracy for the pictures that have no high contrast. For instance, these techniques could mistake the eyebrows of eyes (Feng and Yuen, 1998). Therefore, multiple eye

detection approaches can be incorporated with the aim of increasing the performance of the eye recognition technique. Generally, the dimensions of the eyes and the distance between them are utmost incorporated with the other facial features to establish face recognition. However, it might be advantageous to treat the eye geometry modality as an individual biometric approach to underpin the performance of multibiometric.

2.6.1.4 Ear Geometry

Ear approach measures the shape of the ear and the structure of the cartilaginous tissue of the pinna. Generally, the ear recognition technology depends on matching the distance of prominent points on the pinna from a landmark location on the ear (Jain et al., 2004). Ear biometric technology is viable as the ear anatomy is unique to each person and features based on measurements of that anatomy are comparable over time. That is, each individual has his own ear anatomy that is significantly different from others. Ear biometrics is passive in nature as in face, where it does not require the active participation of the human being (Yan and Bowyer, 2005). It has been demonstrated that the distinctive features of the ear geometry are quite constant over time (Pflug and Busch, 2012). Furthermore, the ear features can be recognised from a distance, and they are not influenced by specific factors, such as lighting and aging (Abaza et al., 2013). Despite the appropriateness of the ear for achieving robust authentication, there are apparently no commercial launches regarding ear geometry technologies. For non-invasive applications, the specifications of the front camera on smart phones may be exploited to achieve transparent ear authentication where the samples can be collected in an unobtrusive manner during a phone call interaction (Clarke, 2011).

On the whole, recognition performance of traditional ear biometric techniques is high, and could reflect encouraging indications towards the initiatives of transparent ear authentication. The common ear recognition performance for specific approaches according to Yan and Bowyer (2005) are tabulated in Table 2.2 as follows:

Table 2. 2 Common Ear Recognition Performance (Yan and Bowyer, 2005)

Approach	Ear Image	Ear Database	Recognition Rate
LABSSFEM	2D	77 (training), 77 (test), USTB ear database	85%
Neural Networks	2D	84 (training), 28 (validation), 56 (test)	93%
Force Field Transformation	2D	252 (test) XM2VTS face database	99.2%
PCA	2D	197 (training), 88 (registrant) ND Human ID database	71.6%
Moment Invariants	2D	120 (training), 60 (test) USTB ear database	96%
Local Surface Patch	2D	10 (training), 10 (test)	100%
Two-step ICP	3D	30 (training), 30 (test)	93.3%
Improved ICP	3D	302 (training), 302 (test) ND Human ID database	98.8%

2.6.1.5 Speaker Recognition

Speaker biometric is the unique representation of the traits which make up a user's voice. The different physical components of a human mouth and throat produce a distinctive sound that can be analysed, measured and stored, and it is well-known as a voice print. Generally, speaker approach recognises the identity of human beings by the distinctive behavioural characteristics of their voice, such as word frequency, the way of speech, pronunciation and accent. Nonetheless, it also can identify people on occasion via physiological traits, such as lips, mouth, nose, glottal folds and larynx (Woodward et al., 2003). Speaker verification approach can be also termed voice recognition or voice authentication. However, it is noteworthy to

differentiate speaker recognition from speech recognition in which the concentration is on what is being said rather than the way of saying (Nanavati et al., 2002). In particular, speaker recognition can be developed in many practical telephony and mobile applications, and theoretically it could operate in the background without forcing individuals to go through a separate verification or identification process in order to provide more usable solutions. On the whole, speaker recognition can operate in two modes which are text-dependant (static) and text-independent (dynamic). In the text-dependant mode, the individual speaks a predefined phrase or given number(s), while the spoken input is free in the text-independent mode. Whilst both of them are viable, the text-dependent mode arguably can offer lower error rates, but with higher intrusiveness (Woodward et al., 2003).

In the conventional methods of speaker recognition, a sample of speech is recorded and analysed as a part of a registration phase (Campbell, 1997). Subsequently, the voice biometric features are extracted by using a sophisticated feature extraction algorithm. The biometric features are then stored as a template in a secure manner. At the authentication phase, a new sample of the same user's voice is recorded and analysed in the same way as above. If the result of calculating the features on verification matches the result obtained during the registration, then the identity is genuine. In order to overcome some security issues, such as ensuring that the pre-recorded voice samples of a person is not replayed to achieve verification, a liveliness detection process can be added to the authentication process in which the caller is asked to repeat sequence of numbers or a random phrase (Toth, 2005).

2.6.1.6 Keystroke Analysis or Dynamics

Keystroke analysis utilises the way in which a user types particular patterns on a keyboard or keypad to collect certain distinctive characteristics, and then to verify whether this user is legitimate or illegitimate. The distinctive characteristics could include the interval time between releasing and pressing a key, hold time of a key press, and the interval time between two subsequent keystrokes which is called inter-keystroke latency (Clarke, 2011). From the perspective of verification, there are two ways can be applied in the keystroke analysis systems: dynamic and static ways. The dynamic technique is a text independant approach which is dependant upon the assessment of the overall users' typing pattern, such as the speed of typing. The static approach is a text-dependant method that means a persons' typing pattern can be examined when they type a pre-specified phrase or word (Banerjee and Woodard, 2012).

Numerous papers have explained that the distinctive actions of keystroke dynamics seem to be insufficient for user authentication (AK et al., 2007). What is more, using the keystroke analysis as a unimodal biometric authentication can be regarded unreliable (Jain et al., 2007). The rate of adopting this technique is also relatively slow (Jain et al., 2007, Clarke, 2011). However, keystroke analysis authentication is deemed quite a convenient modality for multibiometrics and transparent authentication owing to its flexibility in terms of user-friendliness and non-intrusiveness. Moreover, the cost of deploying this technology is very low as there is no need for an additional hardware (Alves et al., 2014). Furthermore, keystroke dynamics approach can be performed as an auxiliary verification mechanism with a view to escalating the level of security. For instance, the Bank of Ireland implemented

the keystroke analysis verification as a second factor in order to enhance the security of the Internet banking services in 2005 (Usman and Shah, 2013).

2.6.1.7 Behavioural Profiling

Behavioural profiling or service usage can recognise an individual depending on the manner of communication patterns with a particular service or device, such as web applications and personal computers (Clarke, 2011). For example, it can build a behavioural profiling for a user who utilises web applications to determine certain attributes, such as duration, access time, date, location, and the sequence of actions. In addition to this, there is a possibility to distinguish the type of applications through tracking the websites that are visited. The performance of generating the initial behavioural template from the profiling attributes is highly likely to be poor. Nonetheless, the user verification via behavioural profile could become robust when communicating with a device or a service on a daily basis; thereby, a consistent profiling actions might be constructed during the period of time that is spent regularly on browsing Facebook every evening or answering emails every morning (Sultana et al., 2014).

Yampolskiy (2008) claimed that there is a possibility to create various behavioural user profiles depending on a particular software interaction to verify whether the same user is interacting with that software environment or not. One example of that is a “web browsing behaviour” that can generate acceptable personal profile identifiers via monitoring a set of events during the user interaction with the online web application, such as using the web application at certain times, typing specific keywords and classifying the web browser type. Another example is an “operating system interaction behaviour” that can generate a behavioural user profile to store

some user behaviours at the time of performing some duties. With Windows operating system, the number of opened windows, the transition time between windows, and the number of the written words in the window title can be taken into account to build different behavioural templates (Yampolskiy, 2008).

According to Tian et al. (2010), a web behavioural profile for user identification could be built by summarising information upon user behaviours and storing them in a database. This information can be gathered and accomplished in either an implicit or explicit way. The implicit information is established through analysing user activities via specific statistical approaches or via data mining. In contrast, the explicit information can be assembled from users through the enrolment stages or through participatory-based studies, such as a user name, address and phone number. In addition, this information can be explicitly collected through user hobbies, such as the number of the user's visits to the favourite websites, the amount of money that spent on an online purchasing (Tian et al., 2010). As such, behavioural profiling can be considered an effective approach for the non-intrusive and continuous verification where it has been applied by some commercial companies for detecting a fraud on credit card and mobile calling devices. Within these technologies, research has shown that the detection rates are more than 90% with low rates of false alarm which may be up to 3% (Stormann, 1997, Clarke, 2011).

To sum up, it is apparent from the above discussion that some transparent biometric approaches cab be taken into consideration for improving cloud-based storage technology. Table 2.3 shows the strengths and weaknesses of transparent biometric approaches according to some distinctive biometric characteristics (i.e. permanence, performance and acceptability) which can be arguably crucial in the context of this

work. Whilst some physical transparent biometric modalities can be superior in regard of permanence and performance, a behavioural transparent biometric technique can have higher acceptability (Jain et al., 2004, Clarke, 2011).

Table 2. 3 A Brief Comparison of Some Transparent Biometric Approaches (Jain et al., 2004)

Biometric Approach	Permanence	Performance	Acceptability
Fingerprint	High	Very High	Medium
Face	Medium	High	High
Keystroke Dynamics	Low	Low	Medium
Speaker	Low	Low	High
Behavioural Profiling	Low	Low	High

On the other hand, a transparent biometric approach still needs to store some biometric features somewhere – thus potential attacks can target the storage component to hack the biometric system. With a view to coping with this vulnerability, research has stepped forward further by evolving the field of biometric cryptography to eliminate the need for storing such sensitive information.

2.7 Biometric Cryptography

With the spread of data communication across the Internet, and the storage of important information through the open networks worldwide, cryptography is increasingly becoming an important if not essential pillar of security. Cryptographic algorithms, such as AES and RSA are being utilised for assuring the authenticity and secrecy of information (Soutar et al., 1999). However, the security of these algorithms depends upon the presumption that the secret keys of the encryption/decryption process are known only to the authentic user (Nandakumar et al., 2007). On the whole, cryptography has no ability to determine whether the person is legitimate or illegitimate (Soutar et al., 1999). In addition, maintaining the secrecy

of these keys is the main challenge in practical cryptosystems (Chang, 2012). Typically, the keys are securely stored in electronic storage, and often protected by password authentication technique. However, passwords can be easily forgotten, stolen, lost, or guessed using social engineering and dictionary attacks (Clarke, 2011). This results in exposing the privacy and confidentiality of the encrypted files. Biometrics can be more secure and usable than password authentication since biometric features could not be forgotten or lost as well as they are quite difficult to be forged or shared easily (Jain et al., 2007). Biometric systems, therefore, can afford a natural and robust solution to the problem of password authentication in cryptosystems (Nandakumar et al., 2007). However, there are also particular security concerns that could influence biometrics. As the source templates (the core of the biometric) are stored at some location within the system, potential attacks can take place upon the storage unit leading to breach the biometric system (Uludag et al., 2004). Further, given encryption/verification credentials (i.e. biometrics, passwords/keys) being stored somewhere, revoking the hacked ones of them cannot be achieved. Consequently, it is apparent from the above arguments that there are still security issues within the biometric and cryptographic systems.

Many authors since 1994 have researched the ability of applying biometrics within cryptography to establish reliable and usable secret keys for security applications (e.g. authentication and encryption). The core approaches of accomplishing bio-cryptographic keys from biometric modalities are varied within the prior research. For instance, a relatively reasonable number of bits can be constantly extracted out from the fingerprint information of 960,000 bits as a biometric key, such as 128-bit. Alternatively, there is another possibility to bind an external key of 128 bits to this

biometric information of 960,000-bit (Cavoukian and Stoianov, 2007). Other approaches also exploited highly robust biometrics such as fingerprint to liberate or release a stored cryptographic key in a secure manner depending on successful biometric verification (Uludag et al., 2004). Consequently, the ideas of extraction, binding or liberation of a key by using biometrics have led to evolve the research area of Biometric Cryptography (BC).

Biometric cryptography is a collection of evolutionary technologies which securely generate a cryptographic key from the biometric, or integrate a secret key into the biometric on enrolment. Instead of storing the biometric template which may be vulnerable to particular attacks, only helper or public data from biometrics are stored – thus facilitating to revoke the security credentials (biometric templates, secret keys, and passwords) in case of compromise. This data will be utilised to reproduce the cryptographic key on verification. Of course, the public data should not help imposters to obtain any information of biometric template or the key. Biometric cryptography can also release a stored cryptographic key in some location on the basis of successful biometric authentication if this location is protected strongly with robust security controls (Uludag et al., 2004, Rathgeb and Uhl, 2011). In all bio-cryptosystems, the biometric key overall is established during the time of enrolment using the reference biometric features, and afterwards the same key should be offered on the verification phase by the test biometric data.

It worth noting that bio-cryptography is not a cryptographic algorithm, such as AES and RSA. In addition, it is worth noting that the process of bio-cryptography is unlike the process of common key generation process in cryptography (Cavoukian and Stoianov, 2007). Biometric cryptography can establish reliable and usable bio-crypto

keys on the fly to overcome the issues of storing sensitive credentials (i.e. biometrics and secret keys with the latter being secured by poor passwords) (Cavoukian and Stoianov, 2007). On the other hand, bio-cryptosystems are still struggling to produce repeatable and constant biometric keys over time as the biometric features can differ each time. Therefore, the significant technical challenge of bio-cryptography is to reproduce the same bio-crypto key despite the natural variations that exist within the biometric feature vector (Cavoukian and Stoianov, 2007). Generally, bio-cryptosystems fall into three basic systems: biometric key release, biometric key generation and biometric key binding (Uludag et al., 2004). These cryptosystems are discussed in the following subsections:

2.7.1 Biometric Key Release

The key release system consists of biometric subsystem and crypto-subsystem. The biometric verification subsystem authenticates the genuine user in order to gain access into the system, and the crypto-subsystem administrates the secrecy of information based on successful authentication. The objective of biometric key release is to reduce the user inconvenience and to cope with the problems of traditional passwords (Uludag et al., 2004). This approach is directly comparable to the existing password-based approach, in which a user password is utilised to protect the encryption keys. The user password is replaced with a biometric-based approach. As such, genuine users no longer need to memorise difficult passwords and strong passwords that cannot be broken via dictionary attacks. The process of biometric key release is illustrated in Figure 2.5:

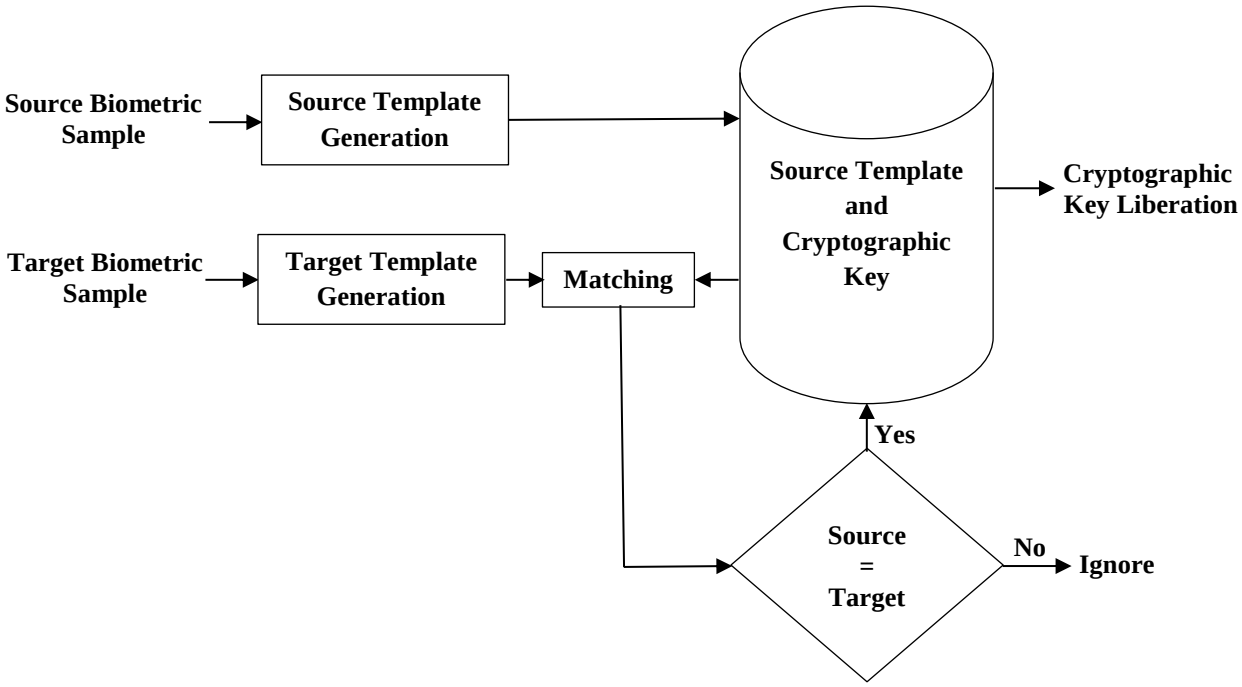


Figure 2. 5 The Idea of Biometric Key Release (adopted from Maltoni et al. 2003)

In the key release system, a biometric sample should be presented to the system when the authentic person needs to gain access to specific resources. When biometric matcher successfully matches between the biometric sample against the reference biometric sample, which is stored during the time of registration, a secret key is liberated to decipher the required resources (Uludag et al., 2004). Biometric key release introduces the good aspect that the cryptographic key will not differ over time as it has no direct relationship with the biometric samples. Although the stored sensitive information (the biometric template and the cryptographic key) are protected by cryptographic solutions, there are a number of security attacks that may impact the key release system negatively. Some of these attacks, which are inherited from the traditional biometric authentication, are discussed as follows (Cavoukian and Stoianov, 2007):

1. Replay Attack: This kind of attack finds a way around the acquisition device via presenting a captured biometric sample into the key release system.

2. **Substitution Attack:** This attack takes place when an imposter can gain an access to the storage unite component of the key release system in which the biometric templates along with secret keys are stored, and thus the biometric template of the authentic person will be overwritten with the biometric template of the imposter. In addition, the secret key can be compromised.
3. **Modification Attack:** Simply this attack means that the feature vector can be altered by an attacker with the purpose of obtaining high matching scores.
4. **Spoofing Attack:** A biometric key release system can be hacked by presenting fake biometric samples. In particular, an adversary impersonates a genuine user identify; thereby, secure resources can be hacked. For instance, an imposter can breach a face authentication system using a picture, or a video recording bearing resemblance to an authentic person (Hadid et al., 2015).
5. **Masquerade Attack:** This category of attack takes place when an adversary can illegitimately obtain biometric sample, and it widely associated with fingerprint and palm print readers. That is, the oils from sweat glands in the skin and residue from touching surfaces will leave a latent print on the surface of the biometric readers. Therefore, it is demonstrated that biometric templates can be typically generated through reactivating these latent prints into readable prints by using a range of techniques including powder, or placing a plastic bag of warm water over the print (Roberts, 2007).

2.7.2 Biometric Key Generation

The concept of this approach is presented to overcome the security issues associated with biometric key release in terms of storing both the biometric template and the encryption key that can lead to malicious attacks. This approach usually but

not necessarily derives public or helper data from the source biometric sample at the enrolment stage (Rathgeb and Uhl, 2011). So as to reconstruct the biometric key, this public data is stored in the storage unit of the key generation system. Afterwards, the bio-crypto key would be generated from the helper data and a live biometric sample depending on the successful verification process (Rathgeb and Uhl, 2011). The verification process is varied in literature, where this process can be accomplished by hash function, file decryption or Hamming distance-based specific threshold. Of course, the stored public data will not help imposters to leak the original biometric template. Thus, neither the biometric data nor the cryptographic key would be stored at some location. Helper data could be a hash value or a vector which indicates the most consistent locations on a template. (Juels and Wattenberg, 1999, Janbandhu and Siyal, 2001, Kanade et al., 2008). On the other hand, it is worth noting that there are other different approaches that can generate keys from biometric without deriving helper data. Figure 2.6 shows the process of biometric key generation by helper data as below:

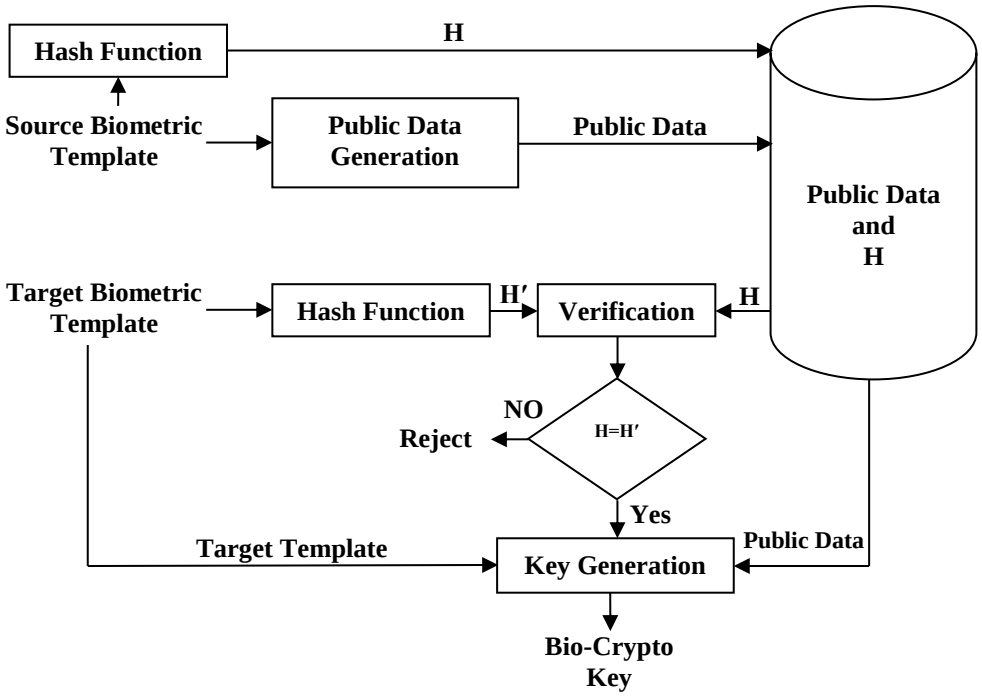


Figure 2. 6 The Idea of Biometric Key Generation by Helper Data (adopted from Rathgeb and Uhl 2011)

2.7.3 Biometric Key Binding

The idea of biometric key binding was firstly introduced by Tomko et al. (1996) to cope with the negative aspects of biometric key generation in which the key is not extracted from biometrics itself. In the key binding approach, an outer cryptographic key is tightly bound with the reference biometric template during the registration phase leading to a latch construction which is stored as public data (Rathgeb and Uhl, 2011). Of course, this latch cannot be broken by imposters to leak the original template as it is constructed via complex mathematical computations (Nandakumar et al., 2007). Additionally, the latch should be built by an irreversible operation, such as XOR bitwise operation. As such, the target biometric template of the valid user will be used to unlock the stored latch (public data) at the time of verification. The unbound key will be utilized for cryptographic goals when the matching process

between the source and query samples is successful (Rathgeb and Uhl, 2011). Figure 2.7 illustrates the process of biometric key binding as follows:

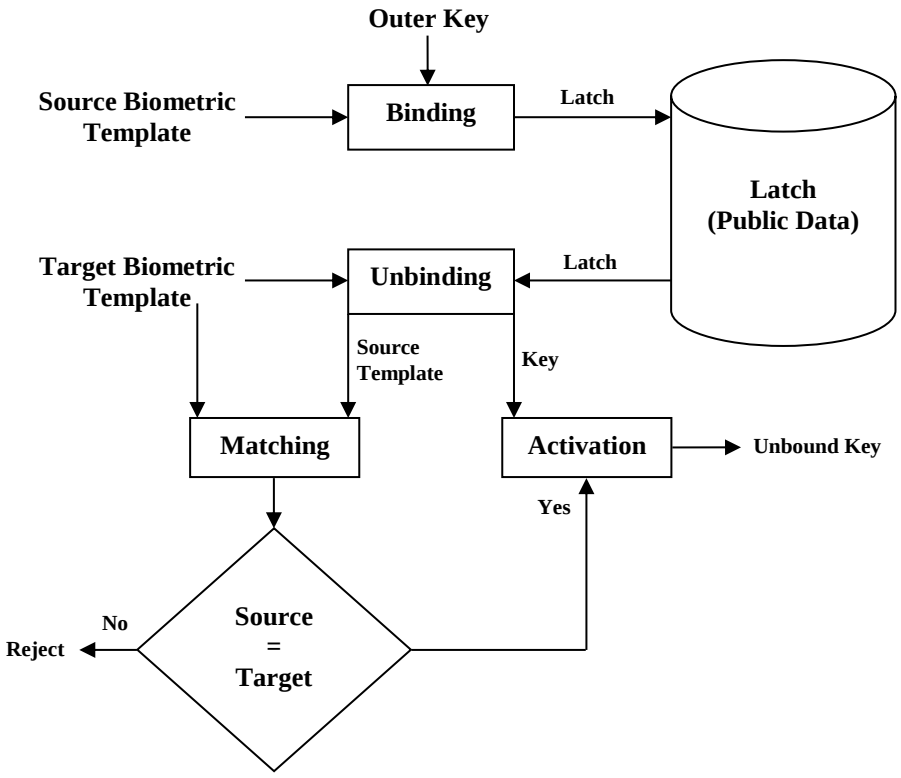


Figure 2. 7 The Idea of Biometric Key Binding (adopted from Rathgeb and Uhl 2011)

2.8 Bio-Cryptosystem Requirements

The objective of the bio-cryptosystems is to provide a mechanism for key production by using a biometric, such as fingerprint, face or voice. Subsequently, this key can be used for cryptographic goals, such as encryption, authentication and data integrity (Soutar et al., 1999). In addition to the requirements of the common biometrics such a universality, uniqueness, and acceptability, there are a number of requirements that should be met first prior to employing the produced key into a specific goal. These requirements are explained as follows (Jain et al., 2008, Kanade et al., 2008):

1. **Revocation:** This characteristic refers to the capability to cancel the key in case of hacking and reissue another one. Whilst this property can be accomplished effectively in the key generation and binding systems, it can be problematic issue in the key release system.
2. **Security Management:** This requirement ensures that insensitive data will be stored within the system in order to overcome the malicious attempts by attackers. This property could be flexibly achieved in the key generation and binding systems, but there are serious challenges to provide robust secure management in the key release system.
3. **Diversity:** This requirement means that various keys should be produced for different applications via the same biometric modalities. This is very important to avoid the cross-matching processes across the databases of that applications. As a result, the user confidentiality will be ensured, and if one application is breached, the other one will not be affected. Arbitrary keys can be easily achieved within the key release and binding systems; however, this could be challenging within the key generation system.
4. **Performance:** This requirement means that the bio-cryptosystem must accomplish a superior classification between legitimate individuals and forgers under which a correct key is established to authentic persons only, and thus enhances the performance. Accordingly, discriminating the biometric features and tolerating their variances should be handled appropriately in order to distinguish between the legitimate and illegitimate users (Soutar et al., 1999). On the whole, biometric variation can be classified into: inter-class and intra-class variations. The former is the variability amongst multiple subjects/users; however, the latter is variance of a single user. Thus, there is

a desirability to raise up the inter-class variations and raise down the intra-class variations for better performance. Biometric variations can arise owing to the weaknesses of acquisition devices and the inherent differences in the biometrics (e.g. aging, poses, and expressions of the face modality). Environmental circumstances, such as illumination can also lead to biometric variabilities. The biometric variances can be reduced by pre-processing methods and/or error correction codes methods which are discussed in the following subsection.

2.8.1 Dealing with Intra-Person Variations

Due to the biometric variances, the effectiveness of the bio-cryptosystem is degraded. Generally, there are two well-known methods in the literature for treating these variations. These methods are explained as follows:

1. **Pre-processing:** This aims to enhance the sample characteristics at the lowest level of abstraction where it either erase unwanted distortions from a sample, or ameliorate some features relevant for further processing and analysis task (Krig, 2014). In particular, pre-processing prepares the biometric sample for feature extraction, and can include the sample size alignment, normalization and noise reduction.
2. **Error Correction Codes:** Ensuring the access of correct data over a communication channel is a challenging task. Data which are specifically collected from biometric modalities are inconsistent over time, where one or more bits could be changed (Wahdan et al., 2013). With the quick developments in technology, the correction of transmitted data become a

more problematic issue. In order to deal with the accidental errors over communication networks, various error correction approaches are proposed for determining if the received data is correct or incorrect without having a copy of the original message. These approaches are depending upon the concept of data redundancy (Kanade et al., 2009b, Shannon and Weaver, 1949). Error correction approaches seek to insert additional redundant bits after converting data into a number of 1's and 0's with the purpose of detecting the bits that corrupted during the transit.

There are two categories of errors within data transmission over network channels; these are single and burst errors. Single bit error means that only single bit has been changed, and it is likely to occur in serial transmission as the error should has a very short duration, but it also could be taken place in parallel transmission (Wahdan et al., 2013). This type of errors is illustrated in Figure 2.8 as below:

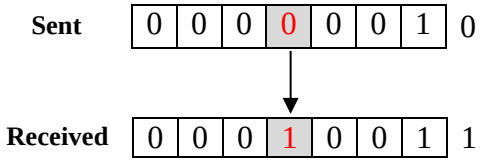


Figure 2. 8 Single Error Category (Wahdan et al., 2013)

However, burst error means that two or more bits have been flipped. Whilst the incorrect bits do not necessarily happen in a consecutive order, the length of the burst error ranges from the first bit to the last bit, and it could include some bits in between which are correct (Wahdan et al., 2013). This type of errors is illustrated in Figure 2.9 as follows:

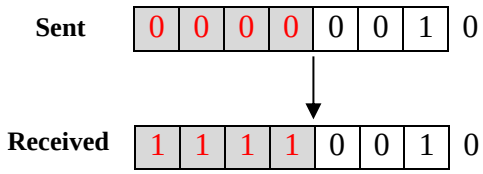


Figure 2. 9 Burst Error Category (Wahdan et al., 2013)

Similarly, the variance in a biometric feature vectors is analogous to corrupt in the transmission of signals, with both single and burst errors possible. The single errors within biometrics can be arisen by the acquisition device impacts. The burst errors, however, can be happened by the inherent noises of the biometrics that can represent the physiological effects (e.g. face aging, and moisture content on fingerprint), behavioural effects (e.g. face poses and expressions, and sample deformation or positioning) and environmental effects (e.g. illumination, ambient temperature, and humidity) (Kanade et al., 2009b). The majority of the biometric features could be stable, but there are also some features will be different. Therefore, Error Correction Codes (ECC) offer good opportunity to correct minor variations in that features. Whilst some of ECC methods can cope with the single errors, such as Reed-Solomon method, the others can deal with the burst errors, such as Hadamard method (Wahdan et al., 2013).

2.8.2 Performance Evaluation of a Bio-Cryptosystem

On the whole, the performance of the bio-cryptosystem is evaluated by set of well-known metrics: FAR, FRR and entropy. As previously documented, FAR and FRR are widely utilised to evaluate the accuracy of biometric systems. While the FAR and FRR evaluation of the key release system is the same as in biometric systems, they can be different within the key generation and binding systems (Rathgeb and Uhl,

2011). The FRR of key generation and binding defines the rate at which wrong keys unsuccessfully produced by the system. That is, the percentage of wrong keys returned to legitimate persons. In contrast, the FAR defines the rate of correct keys returned to illegitimate persons. Compared to the key release systems, key generation and binding overall reflect a noteworthy minimisation in accuracy performance owing to the problematic issue of templates alignment during matching. (Rathgeb and Uhl, 2011).

The concept of biometric entropy determines the distinctiveness of the biometric features (Boulgouris et al., 2009). It provides a measure for the number of possible values a feature vector can take and provides a basis for understanding how well it can withstand a brute force attack. In particular, the entropy evaluation is important for identifying an appropriate biometric for a bio-cryptosystem, and simultaneously can reflect the strength of security (Adler et al., 2006). For instance, the security strength for the key generation system which extracts a biometric key of 50-bit will not be more than 50 bits. The entropy evaluation for a binary face image of 320×320 is 102,400 bits, when all bits are statistically independent. Nevertheless, an authentic user will not be authenticated, if all these bits are inconsistent (Boulgouris et al., 2009).

A direct approach to measure entropy on a random variable X can be done by applying the standard formula (Lim and Yuen, 2016):

$$\mathrm{H(X)} = -\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{Pi}\log_2(\mathrm{Pi}) \quad \ldots\ldots\ldots (1)$$

In the above formula, X is a random variable that represents a set of biometrics features, and pi denotes the occurrence probability of the ith possible value.

2.9 Conclusion

The biometric and cryptographic systems individually are affected by several vulnerabilities and inconveniences. Therefore, the approach of biometric cryptography which combines the principles of biometric approaches and cryptographic algorithms together can deal with these vulnerabilities. Biometric cryptography can offer robust and usable bio-cryptographic keys on the fly to overcome the issues of storing sensitive credentials (biometrics and secret keys) that are secured by poor passwords. At the same time, it prevents users from having to recall, remember those security credentials. Critical key points associated with the bio-cryptosystem implementation should be taken into account. For example, the FAR and FRR evaluation of key generation and binding are different from key release as the key release and the verification processes are independent. As a result, the comparison within key generation and binding is a very challenging process because the key has to be consistent despite the existing biometric variabilities. So as to deal with variances of biometrics, pre-processing methods and error correction methods are utilised to produce the required keys over time.

Transparent biometrics in which the specific biometrics samples are acquired in a non-intrusive fashion can cope with the usability issues that arise from its traditional counterpart. It also can provide an advanced approach through bio-cryptography for protecting information assets in a reliable way. Whilst transparent biometric modalities offer the opportunity to improve upon the usability, the high feature vector variability that is highly likely to result could be a key limiting factor in their application within bio-cryptography.

Chapter Three: Literature Review of Biometric Cryptography

3.1 Introduction

The necessity for an innovative, convenient, usable, and secure encryption solution for cloud storage has been established. Therefore, this chapter is devoted to posing and debating the prior research of biometric cryptography. As such, it establishes a comparison of biometric encryption schemes with the aim of analysing the extent to which these schemes can be applied in practice with transparent and multimodal biometrics. This chapter will ultimately discover the effective approach to the problems of biometric cryptography, and the reliable means for resolving the issues of cloud storage security.

Biometric cryptography is the art of offering strong biometric keys for cryptographic goals, such as encryption/decryption and authentication, aimed at solving the problems of weak passwords (Kanade et al., 2009a). As shown earlier, the biometric encryption approaches can be divided into three categories: biometric key release, biometric key binding and biometric key generation. Using the biometric cryptography approaches, the biometric keys are offered via releasing keys, generating particular consistent features from the biometric sample, or binding secure random bits with constant biometric features on the basis of a successful verification (Uludag et al., 2004). Whilst some authors stated that releasing a stored secret key at some location should not be counted as a bio-cryptographic approach (Cavoukian and Stoianov, 2007), other studies considered it so due to its capacity in handling the weaknesses of password-based techniques (Mariño et al., 2012). Since

the core of this research aims to develop a novel bio-cryptographic approach for cloud-based storage in a more usable and convenient fashion, the literature review concentrates entirely and thoroughly on these three approaches.

3.2 Literature Review Methodology

With the purpose of analysing the current state of the art within biometric encryption in a thorough and robust manner, a particular methodology is employed in order to present a comprehensive review in terms of the issues, challenges, and available solutions resulting in a gap analysis. Biometric cryptography includes a number of terminologies, such as key release, key generation, key binding, fuzzy commitment, fuzzy extractor, secure sketch and fuzzy vault. Consequently, the methodology of the literature review was to search for the above keywords across a range of different academic databases that provide computing resources, such as ACM, Springer, IEEE, ScienceDirect and Google Scholar.

The art of biometric encryption was firstly proposed by Bodo (1994) to overcome the problems involved in traditional passwords. Since 1994, a number of researchers have suggested approaches that utilise biometric characteristics to produce strong cryptographic keys with the capacity to improve security. Accordingly, the literature review will refer to 46 relevant papers that present the current state of the art and explore the research domain. The review is classified into three fundamental sections as shown in Table 3.1:

Table 3. 1 Classification of Literature Review

#	Section	Number of Papers
1	Biometric Key Release	5
2	Biometric Key Generation	27
3	Biometric Key Binding	14

The review will be presented with a thematic fashion in chronological order, where the conventional and the advanced approaches of biometric encryption which incorporate the characteristics of biometrics and the concepts of cryptography together are covered.

3.3 Biometric Key Release Approaches

The idea of biometric key release represents the exploitation of biometric authentication to release cryptographic keys in a secure fashion. This approach stores secret keys in a location, such as a database and then release them on the basis of successful biometric authentication, where the verification and the release processes are entirely separated (Kalsoom and Ziauddin, 2012). Biometric key release presents a positive aspect that the same cryptographic key will be released over time without any variations. However, there are also a number of weaknesses. One of these weaknesses is the potential attack on the stored biometric data, or on the stored cryptographic keys. Moreover, updating the cryptographic key in case of compromising the biometric template is infeasible since it is already hacked (Uludag et al., 2004).

In the literature, there seems to be some confusion in considering biometric key release as a biometric encryption approach. Some researchers have pointed out that bio-crypto key release should not be misunderstood as a bio-cryptographic approach that stores a secret key in a storage unit, and then releases it on the basis of

successful biometric authentication (Cavoukian and Stoianov, 2007, Rathgeb and Uhl, 2011). On the contrary, other studies have categorised this approach as a biometric cryptosystem as it is one of the primary approaches that solve the problems of weak passwords, and present a convenient manner to prevent the user from having to recall complicated passwords (YEE, 2011, Mariño et al., 2012). Clearly, biometric key release approaches depend on the performance of the biometric recognition subsystem in releasing the cryptographic key. As such, few studies have specifically focused upon this as it fits into the wider body of research that seeks to improve the underlying performance of biometric modalities. Therefore, this section will present only the available papers that deal with the key release idea.

The primary model of biometric key release was introduced by Soutar et al. (1999) who used the fingerprint modality because of its high recognition performance in the biometric authentication area. In a brief review, the source template along with a cryptographic key is stored in a secure storage unit. Subsequently, on authentication, the live fingerprint of the genuine user is verified with the stored the source template. Then, if biometric verification is successful, the cryptographic key will be released. On the whole, the separation between the biometric matching and the key release processes permits the security subsystem to revoke the privileges and the rights of the user. However, the presence of a biometric template for authentication causes specific vulnerabilities.

Other papers presented a conceptual review of the existing systems of biometric template protection – particularly on biometric key release. Specifically, the authors addressed the properties of the key release system by two key points. Firstly, this system needs access to biometric templates to perform the matching process.

Secondly, the key release and the verification processes are wholly independent. At the same time, the writers claimed that these characteristics result in serious security issues. In simple terms, due to the local storage of important data, there are some problems that may occur when designing such system. Among these problems is the capability of adversaries to steal the stored biometric data from one application, and then utilising them in another application. In addition, as the matching process is totally separated from the key release process, the system will be vulnerable to attacks in which match or no match response can be hijacked and masqueraded (Uludag et al., 2004, Rathgeb and Uhl, 2011).

As mentioned previously, the key release system relies upon the accuracy of the recognition subsystem. Therefore, a review by Kalsoom and Ziauddin (2012) discussed the concerns surrounding iris recognition systems. The authors claimed that the investigative effects upon reliable feature extraction would certainly improve the performance of iris-based key release system if they are handled appropriately. Some of these factors include the pupil dilation, the usage of contact lenses and the similarities between the irises of twins, where all these factors can influence the system. On average, iris recognition systems reported a very good recognition rates that ranged from 69% to 100%.

Another contribution to biometric-based key release was published by Karovaliya et al. (2015). Face recognition subsystem by Principal Component Analysis (PCA) and One-Time Password (OTP) authentication were used for Automated Teller Machine (ATM) to improve confidentiality. The recognition by PCA will reduce the forgery risks associated with the theft of smartcards, and OTP will not oblige users to remember their long passwords. The enrolment process is carried out by the authorized

employee in the bank. The employee would acquire some face images of the genuine user by an advanced acquisition device in the bank, and store them in a database. On authentication, once the ATM smartcard is swiped, a face image is acquired via a camera installed upon the ATM, and matched with the stored images in the database. When verification is successful, an OTP code is sent to the recipient's mobile number. So as to authenticate the user, the comparison operation between the forward OTP code from the bank to the user and the backward OTP code from the user to the bank is achieved via a one-way hash function. The transaction will be processed if the intended user has input the right OTP code within three trials. Otherwise, the account will tentatively sign out. One Time Password (OTP) is a conventional authentication technique which can authenticate the legitimate user via creating an OTP code for each session or transaction. This method is employed to cope with the concern of having to remember the long and complex passwords (Clarke, 2011). Therefore, One Time Password (OTP) can be arguably considered as a key release approach.

However, specific criticisms which could affect the proposed system are noticed. One of these is that the loss of mobile/internet connection would not transmit the OPT code to the recipient user, and would interrupt the transaction process. Additionally, there is a rather inconvenient aspect in using the OTP technique, where the user should input different OTP(s) frequently. Another criticism is the lack of experimental works, where there is neither a rigorous security analysis nor accuracy tests in terms of FAR and FRR results that would manifest the ability to debase potential attacks and to make this scheme applicable in such realistic environment.

The approaches of biometric key release are summarised in Table 3.2 as below:

Table 3. 2 Short Summary on Biometric Key Release Contributions

#	Authors	Year	Biometrics	Security	Recognition	Verification	Type
1	Soutar et al.	1999	Fingerprint	Cryptographic Algorithms such as DES	NA	NA	Conceptual
2	Uludag et al.	2004	Fingerprint	–	–	–	Review
3	Rathgeb and Uhl	2011	Fingerprint – Iris – Face	–	–	–	Review
4	Kalsoom and Ziauddin	2012	Iris	–	–	–	Review
5	Karovaliya et al.	2015	Face	One Time Password	PCA	MD5	Prototype

As shown in Table 3.2, there is a lack of technical research in resolving the issues associated with the key release scheme. The majority of the related works concentrated upon the conceptual approaches of biometric key release. Whilst biometric-based key release has the capability of dealing with usability, the privacy of the stored data is considered the greatest obstacle for implementing and adopting these systems.

3.4 Biometric Key Generation Approaches

Numerous contributions have been published in the research area of combining biometric and cryptography to present a biometric key generation approach that securely generates keys for encryption and decryption, authentication and access control purposes. Biometric key generation approaches can be classified as either direct key generation or indirect key generation. Direct key generation can generate biometric keys directly from the source biometric template, and then discard them at the end of the session, such as encryption. For decryption, this approach should regenerate the same biometric key from the test biometric template of the same valid

user (Rathgeb and Uhl, 2011). However, indirect key generation approaches derive helper data from the reference template to generate indirect keys. Of course, for maintaining security, the storage of the helper data should not be useful to forgers to reconstruct the source biometric template. The negative aspect of the key generation scheme overall is the difficulty in regenerating the same key over time with high consistency and complexity (Uludag et al., 2004).

The direct key generation approach is still an open to challenge due to the lack of helper data storage that hinders the generation of a constant key over time. However, there have been specific attempts that contribute to overcome this challenge. The prior proposal of generating biometric keys directly from biometric template was suggested by Bodo (1994) in a German patent; however, there was no practical implementation. Janbandhu and Siyal (2001) corroborated Bodo's proposal and contributed to the generation of a private key by RSA and DSA algorithms through an iris biometric where the iris features are used as auxiliary means for the key generation process. With the purpose of obtaining a highly stable iris template, the authors used an off-the-shelf iris recognition product by Iriscan Company to generate a 512-byte iris code with a very good EER value of 1 in 1.2 million. The generated RSA key was 512-byte whilst the DSA key was 160-bit.

In the same context, Hoque et al. (2008) enhanced the vector quantization method to explore the possibility of generating a biometric key squarely from handwritten signatures. This work was inspired by the idea of Yamazaki and Komatsu (2001) which stated that Vector Quantisation (VQ) is used to remove the variabilities of biometric samples, where the extracted features are partitioned into a specific number of cells, and each one is symbolized by using mean vectors. Subsequently,

the target samples will be compared with those vectors, and the closest vector in the codebook specifies whether the person is legitimate or illegitimate.

The enhancement of the vector quantization method consisted of replacing the codebook with a group of partitions which are induced in the feature subspaces, each subspace being constructed by means of subset of features. The partitions determine the number of cells in those subspaces, where each cell is labelled with a certain identity. On key regeneration, the feature subspaces of the test sample which are symbolized by their own group of partitions are processed individually to regenerate the biometric key by concatenating them. Individuals do not need to introduce their identities to gain access to an encrypted document. The ability of providing a biometric sample that can decrypt the document is adequate evidence of verification. It is supposed that all biometric features were distributed normally.

The analysis and evaluation of this investigation were conducted upon a database that included 144 respondents with 15 samples from each one. In addition, 133 active imposters were engaged to forge particular signatures. On the whole, a biometric key of 32 bits can be regenerated at 35.2% and 5.6% FAR and FRR respectively where the number of partitions was 5. Nonetheless, this biometric key was short, and could possibly be broken by brute force attack. Further, although the FRR result was rather acceptable, the FAR result defeated the system where more than quarter of imposters may be accepted incorrectly.

In another work, Sheng et al. (2008) suggested a direct key generation schema from the statistical features of handwritten signatures, such as pen-down time, pen-up time, the overall duration of the signature and the number of strokes via specific statistical approaches. The researchers claimed that the methods of clustering

biometric data, such as fuzzy clustering and vector quantization are applied to limited and predefined code words or clusters, and are likely to be influenced negatively by suboptimal features. Consequently, fuzzy clustering was enhanced by applying genetic algorithm to identify the suitable and close-optimal features in the training handwritten signature samples. Accordingly, the stability of each feature, whether it is a subset or single, is quantified for each individual. At end, the biometric key is reliably generated when the most stable features are chosen. In order to conduct the analysis and the experimental results, 7430 handwritten signature samples of 359 participants were gathered in public trials of an automatic signature authentication system. Overall, the authors generated a biometric key of 20 bits at 0% FAR and 14.5% FRR recognition performance. However, the produced bio-crypto key was quite short that could be easily broken by brute force attack. In addition, there was no evaluation concerning the entropy of the generated biometric key.

The idea of analysing the biometric features comprehensively with the purpose of investigating the most consistent features that could reliably generate a constant key from biometrics squarely was proposed by Atah and Howells (2009). The authors explored the appropriate features of a speaker modality, such as maximum power spectral density and maximum amplitude by using the built-in facilities of a Microsoft feature extractor and the 'wavread' function in Matlab. The researchers claimed that the unsuitable normalisation methods for obtaining common distributions were feasible on a range of features. Consequently, in order to diminish the within-user (intra-class) variances, the resultant consistent features were normalised empirically by a min-max normalisation method to become limited into a common scale between 0 and 1. Accordingly, the normalised features were multiplied by a constant variable

to become accurate in a decimal system, then transformed into binary values via a quantization criterion.

The public database of VALID which included 106 participants was adopted to test this work. Each user had five sample recordings of the uttering “Joe took father’s green shoe bench out” in specific noisy environments. This work showed that the accuracy of reproducing the same direct key over time was 65%. However, this research did not demonstrate the recognition performance respecting the FAR and FRR results. In addition, the effective bit space was not provided.

Regarding the research area of multibiometrics, a number of direct key generation approaches from multiple biometrics have been proposed. Among these approaches, Jagadeesan and Duraiswamy (2010) combined fingerprint and iris biometrics, while Abuguba et al. (2015) fused face and the iris modalities. Both studies utilized off-the-shelf Daugman iris technology to generate the feature space of the iris modal; however, morphological transform and PCA were applied to the fingerprint and face biometrics respectively to extract their multidimensional features efficiently. For both contributions, the researchers sought to randomise the feature spaces for security purposes. Jagadeesan and Duraiswamy (2010) concatenated the fingerprint and iris vectors after a number of scrambling steps to obtain a uniform multibiometric template. Accordingly, a 256-bit bio-crypto key is constructed through taking the modulus two for each value over the constructive template.

On the other hand, the face and iris fusion along with the key generation are simultaneously applied in the research of Abuguba et al. (2015). That is, the iris is represented by 256 integer values whilst the face by 256 bits. As a result, the i-th bit of the key is the summation of all the zero values in the binary system of i-th

normalized iris values modulus two if the i-th bit of face local binary pattern is zero. In contrast, the i-th bit of the key is the summation of all the one values in the binary system of i-th normalized iris values modulus two if the i-th bit of face local binary pattern is one.

For both studies, the experiments were carried out by using public databases for the fingerprint, the CASIA database for the iris and the ROL database for the face. Both contributions succeeded in generating a 256-bit biometric key from the multibiometric templates. Abuguba et al. (2015) demonstrated an FRR result of 12.51% which is a rather unacceptable value. This work does not state the accompanying error rate in terms of FAR and FRR. Jagadeesan and Duraiswamy (2010), however, did not perform empirical studies illustrating the recognition performance of FAR and FRR. In addition, although the contributors claimed that the biometric samples were pre-processed well by noise reduction algorithms to eliminate the emerging variances, there was no evaluation of intra-class variations to demonstrate the stability of reproducing the same biometric key over time. The authors also did not assess the effective bitlength of the bio-crypto key using the biometric entropy to show how strong it was vis-à-vis brute force attacks.

In a follow-up research, Balakumar and Venkatesan (2011) replicated the proposed system by Jagadeesan and Duraiswamy (2010) on proprietary databases that included 100 participants. The authors succeeded in regenerating a biometric key of 256 bits at an overall FAR and FRR of 0.2351% and 85.07% respectively. However, this accuracy figure of FRR reflects that the proposed system by Balakumar and Venkatesan (2011) may not recognise one third of the legitimate users.

On the other hand, indirect biometric key generation approaches are more tolerable to errors in the context that the generation or the extraction should be the same even whether the query biometric template has small variations. Among these approaches, fuzzy commitment can be considered as an indirect key generation that utilises error correction codes to eradicate the biometric variabilities. In this context, the term fuzzy refers to the fact that the source and the test templates are near or close to each other, and not entirely equal due to intra-class variances (Juels and Wattenberg, 1999). Indirect key generation approaches are also proposed as a fuzzy extractor or secure sketch approach. The former generates public data from the source biometric sample that can create the key via the live sample of the same user. The latter, however, uses the helper data to construct fuzzy sketches from the reference and the live biometric samples for verification purposes (Li et al., 2006).

There has been significant amount of research upon indirect biometric key generation. Nevertheless, their categorisation into specific areas is not easy because particular approaches can be equally applied to key binding and key generation, such as fuzzy commitment (Cavoukian and Stoianov, 2007). Consequently, a particular characteristic is required in order to decide whether these approaches are key binding or generation. Conceivably, the manner of producing the key can be adopted as a characteristic to differentiate between the key binging and generation approaches. As such, if the approach binds an outer key with the biometric signals, then this key binding; otherwise it is key generation. The review of indirect key generation approaches can be classified into: fuzzy commitment, fuzzy generator, and fuzzy or secure sketch. These approaches will be discussed in the following subsections:

3.4.1 Fuzzy Commitment Approaches

On the whole, the fuzzy commitment approach handles the noises that exist in biometric samples via the methods of error correction codes. Hamming distance of specific threshold between the source and the target biometric templates decides whether or not these belong to the claimed user. Various error correction codes can be used to eradicate the biometric variations, such as Reed-Solomon (RS), Bose-Chaudhuri-Hocquenghem (BCH), Low-density parity check (LDPC) and Hadamard (Wahdan et al., 2013).

The fuzzy commitment approach was firstly introduced by Juels and Wattenberg (1999). In this approach, a biometric template which is represented as a binary string *b*, is added to a random codeword *cw* via an error correction approach in order to calculate the commitment function c by:

$$c = cw + b \quad \ldots\ldots\ldots (2)$$

At the same time, the values of *c* and the hash value of *cw* are stored somewhere for authentication purposes. On decommitment, the codeword *cw´* is regenerated via:

$$cw´ = b´ - c \quad \ldots\ldots\ldots (3)$$

where *b´* is the binary string of the test biometric template, and the verification is successful if the hash value of *cw´* equals the stored hash value of *cw*. Concerning the codewords with a minimum Hamming distance *d*, the authentication cannot be unsuccessful when the Hamming distance between *b* and *b´* is lower than or equal to *d/2*. Noticeably, the commitment term refers to the fact that the error correction codes are committed to overcome the biometric variabilities on authentication. Figure 3.1 shows the block diagram of the fuzzy commitment idea:

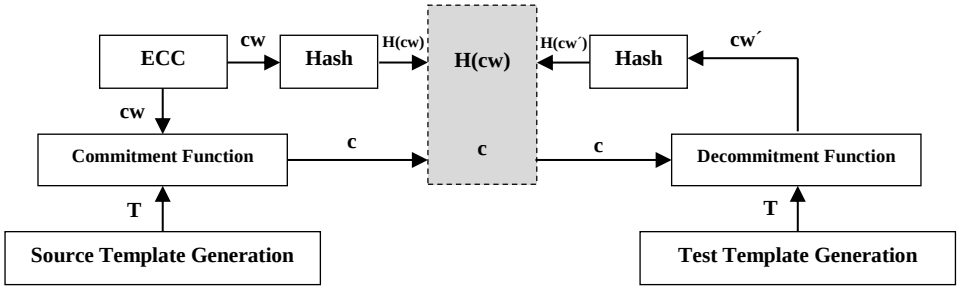


Figure 3. 1 Block Diagram of Fuzzy Commitment Idea (modified from Juels and Wattenberg 1999)

In a different review, Kanade et al. (2008) suggested a fuzzy commitment framework for smartcards by using the iris modality and a combination of Hadamard and RS codes. Of course, Hadamard codes can erase the inherited noises by the capture device from biometrics while RS can adjust the inherent changes in biometrics, such as placing the biometrics on the acquisition devise differently. In this research, a scrambling approach by using a password shuffles the iris codes to maximise the separability between legitimate individuals and forgers depending on a Hamming distance metric. In addition to ECC, a procedure of inserting and truncating extra zeros is utilised among the iris codes to raise up the error correction ability to over 25%. Whilst XOR bitwise operation was exploited to increase the randomisation aspect of the iris code, the hush function was relied upon for verifying the authentic user. This work was performed on a public database Iris Challenge Evaluations (ICE) that included 124 users. The accuracy results were 1.04% FRR and 0.055% FAR, and the bitlength of the regenerated key was 198 bits with an 83-bit estimated entropy. The authors have recognized the difference between the bitlength of the regenerated key and its effective bitlength in combating brute force attacks where they evaluated the key entropy based on the fact that not all values had an equal contribution.

In the forth work of Kanade et al. (2009b) a weighted error correction schema was proposed by combining the RS and Hadamard codes. This combination aims to eliminate the biometric variances from the right and left iris modalities at the time of verification. That is, the Hadamard decodes can correct a specific capacity of errors, and these errors will be decoded wrongly if they are more than this capacity. As a result, the Hadamard decodes will input to the RS decodes to fix the errors which are incorrectly decoded via Hadamard error correction. Initially, the valid person will select a password to randomise the symmetric codes of RS by using scrambling approach for security purposes. At the same time, this password will reinforce the revocation requirement, where the hacked one can be simply cancelled to select another password. Accordingly, the resultant asymmetric codes of RS are encoded by the Hadamard codes to rise up the correction capacity, and this will obtain the pseudo code S. For the right iris code I_1, a zero insertion and truncation manner is employed symmetrically, on registration and verification, to originate the modified version II_1, and the latter is concatenated with the left iris code I_2. At end, the iris codes concatenation (II_1 and I_2) is xored with the pseudo code S to produce the lock code I_{Lock}.

Having established Hadamard error correction decoding at authentication; the right iris code produces a greater number of correct decodes than the left iris code. If the number of incorrect blocks is lower than or equal to ts, the error correction capacity of the RS, they are correctly decoded by the RS decoder to generate the bio-crypto key. While the number of correct blocks via the Hadamard method can be noticed as a similarity score, the error correction capacity of the RS (ts) can be considered as a threshold dependant classifier. Therefore, the variable error correction represents a

weighting approach for features consolidation, where the right iris code has a higher weight than the left iris code. The more the error is corrected for the right iris code the less the valid users are rejected by the proposed system. However, the less the error is corrected for the left iris code, the more the valid users are rejected. The composite impact of both iris modalities leads to enhance the system accuracy.

The experimentation of the proposed system was implemented on the public database Iris Challenge Evaluations (ICE) that included 120 respondents. The bitlength of the reconstruct bio-crypto key was 147-bit and the recognition performance was 0.18 FRR and 0% FAR. However, selecting weak password will affect the system negatively when an attacker can guess the poor password. As a result, sensitive secure information will be hacked with the purpose of capturing the cryptographic key. Further, there is a particular criticism presented by Stoianov (2010) that the procedure of extra zeros insertion and truncation can be broken by a hacker. The author claimed that the locations of the extra zeros within the iris code are already known; therefore, an adversary could analyse them to regenerate the biometric key within a reasonable time.

In further attempts at fuzzy commitment exploitation to remove biometric variances, Teoh and Kim (2007), Sutcu et al. (2008) and Wahdan et al. (2013) employed the error correction coding to assist in managing certain challenges. One of these challenges was how to deal with the biometrics of multiple dimensional features, such as face and fingerprint. The 2D representation of minutia features, fingerprint features, is large and complicated; therefore, the variations between the constructive templates will be high, and this will influence Hamming error calculation negatively (Sutcu et al., 2008). Teoh and Kim (2007) extracted the biometric features of

fingerprint by Gabor transform, after which these features are discretised/binarized into binary values using a randomized dynamic quantization transform. This transformation includes an irreversible process depending on random numbers that are generated via a user-specific token. Of course, this token should be stored on a secure device to generate the same random numbers at authentication. Accordingly, 375-bit of binary distinct features is obtained with fairly uniform randomisation. Afterwards, with the purpose of removing the biometric variations, symmetrical RS codes are xored with those binary features to originate the locked codes. The researchers used the public fingerprint database DB1 from FVC2002 website which included 100 participants with 8 samples for each participant. The recognition performance of FAR and FRR were 0% and 0.9% respectively, with a biometric key of 375 bits. However, there was no evaluation with regard to the effective key entropy.

Another study in the same context was conducted by Sutcu et al. (2008) who suggested probabilistic scheme based on statistic methods, called user-specific cuboid. This scheme extracts the minutiae points of fingerprint and their orientations to correspond with the LDPC codes for the binary symmetric channel (i.e. the communication channel for binary information). A user-specific cuboid scheme calculates the number of variations among the minutiae inside and outside the cuboids, and accordingly uses Bernoulli statistical approach to produce the binary values. The experiments of this research were conducted on a proprietary database that included 1035 respondents with 15 fingerprint samples from each respondent. This research reported accuracy results of 0.11% FRR and 1.0019% FAR. The statistical analysis illustrated that the scheme of user-specific cuboid can efficiently reduce the errors of binary symmetric channel via LDPC codes. In terms of good

aspects, the work of Sutcu et al. (2008) evaluated the intra-user variations by using a histogram approach. Nonetheless, the key bitlength was 30 bits which is fairly short and could be broken by brute force attack.

Whilst the previous two studies concentrated upon tackling the variances of unimodal biometrics, another investigative challenge involving the assistance of error correction codes is how to eradicate the variations of multi biometric feature space. Hypothetically, Wahdan et al. (2013) used a Reed-Solomon error correction scheme individually upon each biometric feature to solve this issue. In terms of feature extraction, Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT) and cepstrum analysis methods were used to extract the features from the iris, face and voice biometrics respectively. Afterward, a binarization approach was used to transform the features of each modal into a binary string using the mean value of the resultant features. Accordingly, the multiple binary features are consolidated depending on the feature fusion level via an Advanced Encryption Standard (AES) algorithm. Apparently, the multi biometric features are input into AES as a plaintext, and then an outer cryptographic key is used to concurrently encrypt and combine those features in a secure fashion. Finally, the resultant binary features are used as a bio-crypto key for encryption and decryption purposes. The proposed system was carried out upon a proprietary database with 20 volunteers. The bitlength key generation was 192-bit with an overall recognition performance of 99.83%. Nonetheless, the drawback of this study was the lack of a security analysis. Furthermore, when outer cryptographic keys are used, the system may be hacked if the forger can attack the stored password at some location. What is more, there was

no evaluation concerning biometric variances – specifically for the biometrics of geometrical or multidimensional features, such as face modality.

3.4.2 Fuzzy Extractor or Generator Approaches

Generally, the fuzzy extractor approach can benefit from the fuzzy commitment idea in utilising the ECC to remove the biometric variations as well as it extracts helper data from the reference template that contributes to generate the biometric key via the test template of the same user.

The primary paradigm of the fuzzy extractor was suggested by Dodis et al. (2004) to generate a biometric key with the ability to handle biometric variations. On key generation stage, both the bio-crypto key and helper data (the public data) are constructed from the reference template and only helper data are stored in the system to be used at the time of verification. Subsequently, the test biometric template and the public stored data are used to regenerate the key. Of course, the helper data storage must not help hackers to leak any information about the key and the train template. In order to verify the key, fuzzy extractor approaches use an error distance metrics-based specific threshold such as Hamming distance, Edit distance and Set distance to calculate the biometric variances between the source and test templates. Figure 3.2 illustrates the block diagram of the fuzzy extractor approach:

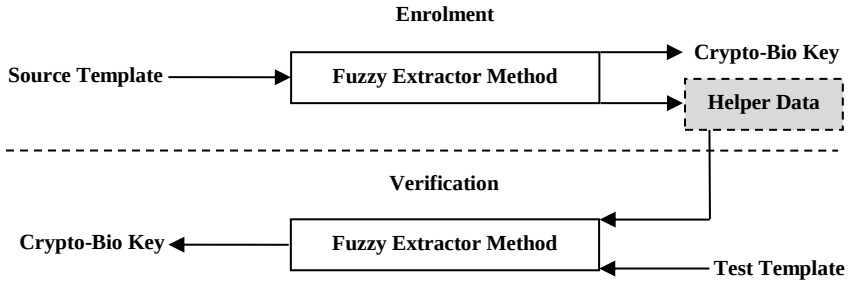


Figure 3. 2 Block Diagram of Fuzzy Extractor Approach (adopted from Li et al. 2015)

In further studies, Li et al. (2012) and Feng and Yuen (2012) claimed that a fuzzy generator cannot be implemented for biometrics, such as fingerprint and face modalities since their features are high-dimensional geometric features. In this situation, distance metrics-based specific threshold cannot calculate the errors, variations, between the biometric templates. As a consequence, Li et al. (2012) used a combination of RS and BCH, BCH, and LDPC codes to determine the optimal method for erasing the variances. Additionally, the probabilistic scheme of a user-specific cuboid by Sutcu et al. (2008) was used appropriately to generate a discriminant binary key from the real-valued features of a fingerprint minutiae triplet. The authors claimed that the use of minutia triplet features does not require the implementation of the tricky alignment process that is applied among the biometric samples to become common versions as these features are fairly stable over time. As aforementioned, the probabilistic scheme of user-specific cuboid statistically calculates the differences between the minutiae inside and outside the cuboids among a number of fingerprint samples of the same user. However, Li et al. (2012) reduce the intra-user variabilities further by determining the biometric variances through these calculated differences, where their smallest values are reserved and indexed as a robust vector *rv* since they represent the most reliable regions on the fingerprint template. Moreover, the Liner Discriminant Analysis (LDA) method

contributes to eradicate the inter-user variabilities by its resultant Dimension Reduction Matrix *DRM*. Accordingly, mean and Bernoulli statistical methods where used to define a discretised vector *Dv* and to produce a fixed binary key *Bk*. Of course, the helper data of *rv*, *DRM*, and *Dv* are stored somewhere to be used in regenerating the same binary key during the authentication phase, using the test fingerprint template.

For security purposes, the XOR operation is performed between the fixed binary key *Bk* and the resultant codeword *cw* from the error correction codes to obtain the locked data *Ld*. Simultaneously, the *cw* is secured through applying the hash function *H(cw)*, and will be stored together with *Ld*. At the verification time, a refresh fingerprint sample is presented to generate fixed binary key *Bk´* by using the stored helper data. As such, the XOR operation is applied between *Bk´* and the stored *Ld* to obtain *Ld´*. Finally, the error correction codes are carried out on *Ld´* to produce codeword *cw´*. The authentication process will be successful if *Bk* and *Bk´* are from the same genuine user, and within a specific threshold of Hamming distance error.

The experiments of this research were performed by using the public database FVC02 DB2 which included 110 volunteers. The empirical results demonstrated encouraging accuracy results of 0% FAR and 4.85% FRR. However, the bitlength of the generated key was 48 bits which is fairly short, and could be broken by brute force attack. Interestingly, the error correction coding of LDPC outperforms the other methods, BCH and the combination of RS and BCH. Figure 3.3 shows the block diagram of the registration and authentication processes:

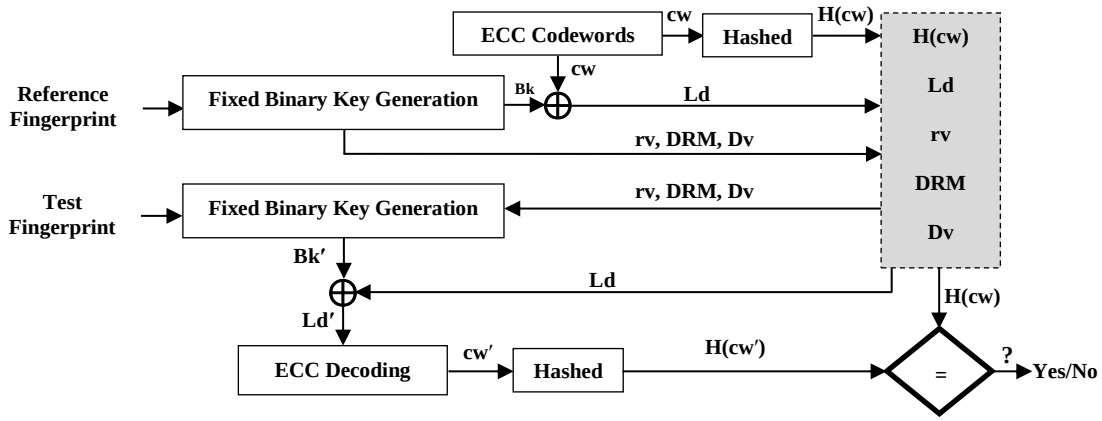


Figure 3. 3 Block Diagram of Registration and Authentication (modified from Li et al. 2012)

In the same context, that of investigating the challenge of multidimensional biometric features, a neural network algorithm was exploited by Feng and Yuen (2012) for identifying the discriminant features from a face modality. In this research, during the time of enrolment the neural perceptron trains the binary values to produce the most consistent bits and to set public data that assists in regenerating the same discriminant features later if necessary. Of course, the stored data should not help adversaries to recover any important information for the original template. At the verification stage, the neural network transforms the test face sample into the discriminant binary template by using the stored helper data. Concerning the feature extraction methods, random projection, Eigenface and Fisherface algorithms are used to investigate the best feature vector generation. Afterward, the error correction coding of BCH is appropriately used to eliminate the variances between the generated bio keys. Consequently, the test binary template will be same or near to the trained binary template. The experiments in this research were conducted with 68 participants, and it encouragingly demonstrated that the recognition rate was 96.38%. Moreover, the Fisherface algorithm powerfully generates the face template and outperforms the other methods, random projection and Eigenface. Furthermore,

various BCH codes do not impact the system performance badly. Moreover, the entropy of the produced key was 75 bits. However, the long execution time of system implementation probably may be considered as an inconvenient aspect to end-users.

In other works, aimed at eliminating the variations of 2D dimensional features such as fingerprint minutiae Dodis et al. (2004) and Chang and Roy (2007) proposed fuzzy extractor approaches depending on biometric combination mechanism. Biometric combination was proposed by Dodis et al. (2004) for fuzzy extraction via embedding one metric space into another metric space, which is evaluated well based on error distance functions through subsequent biometric acquisitions, and then stored as helper data. The goal behind this embedding is to maintain relevant values for fuzzy generation. As a consequence, the fuzzy extractors in the source and target templates will be close to each other. Chang and Roy (2007) used PCA along with biometric combination to investigate the most consistent features among the minutiae of fingerprint biometric; therefore, the variations were further diminished. Accordingly, a binarization-based threshold method is applied to features in order to transform them into binary values. From the other perspective, with the purpose of maximising the security and performance, the XOR operation is applied between the binary values and the codes of ECC to generate the crypto-bio key. With regard to analysis and evaluation, the empirical work was performed upon 4000 fingerprints from the database NIST. The generated biometric key ranged between 8 to 10 bits with an accuracy result of FRR = 0.09%. This paper does not illustrate the accompanying error rate in terms of FAR and FRR. Given the weak entropy, the biometric key could be used in the environments that require low level of security, or

where access to the system is protected by other measures (i.e. that limit the opportunity of brute force being feasible).

With regard to the research area of multimodal biometric and without using the ECC methods, Chin et al. (2014) presented a fuzzy generator approach using the fingerprint and palmprint biometrics. At first, once the biometric features are extracted via a number of Gabor transformations, the combination-based feature level is applied by rearranging these features to obtain a multimodal feature template. Subsequently, in order to achieve the revocability requirement for the bio-cryptosystem, a random tiling transform-based specific key is used to generate the irreversible version from the multimodal template. Specifically, random tiling configures random rectangles with dimensions determined by a specific tokenized key. Consequently, the user should choose a new key to reissue another multimodal feature template in case of compromise. In the same context, another security issue is that the imposter could expect the binary bits by combining the 1's and 0's of the highest probabilities. So as to avoid this concern, the equal probable binarization technique splits the multimodal feature template into equal width intervals, and distributes the features equally among them by a statistical distribution approach. Each interval is indexed and labelled with an integer, and its features are transformed into 1 or 0 values based on their label. Finally, the biometric key is composed by concatenating all binary bits.

During authentication, the same steps as those above are applied to generate a test binary key by using the stored helper data, the specific tokenized key of random tailing transform. At that time, the matching process compares between the test

binary key and the source binary key via a Hamming distance metric to decide if the test biometric belong to the claimed user.

300 volunteers were involved to conduct the tests of the proposed system by collecting 8 samples from each volunteer. The recognition performance was roughly 0.05% EER with feature length of 200 dimensions. These results demonstrated the effectiveness of the random tilling and the equal probable binarization methods where they are feasible in multibiometric fuzzy extractor. The only criticism of this research was that, in spite of the pre-processing stage which was applied to the biometric samples by noise reduction algorithms, there was no evaluation of intra-class variations.

In a different review, Chang (2012) proposed dynamic private key generation for public key cryptosystem. The author employed the keystroke recognition depending on the RSA algorithm to resolve the issues of key management. Principally, those issues involve the negative ways of protecting and storing the private key using password verification. The interesting idea behind this research is that a keystroke recognition approach gathers the keystroke features, and accordingly neural network algorithm trains these features to generate the private key. During the time of enrolment, keystroke actions are concurrently gathered by typing the valid password. Afterwards, specific keystroke features are learned by back propagation neural network to produce a target random private key of 2048-bit, which is already generated by RSA. When the learning process is completed, only the parameters of RSA and the back propagation, such as the public key, random integer and weights are stored as helper data in the user's storage unit to reconfigure the neural network again if necessary. Figure 3.4 illustrates the block diagram of the key learning stage:

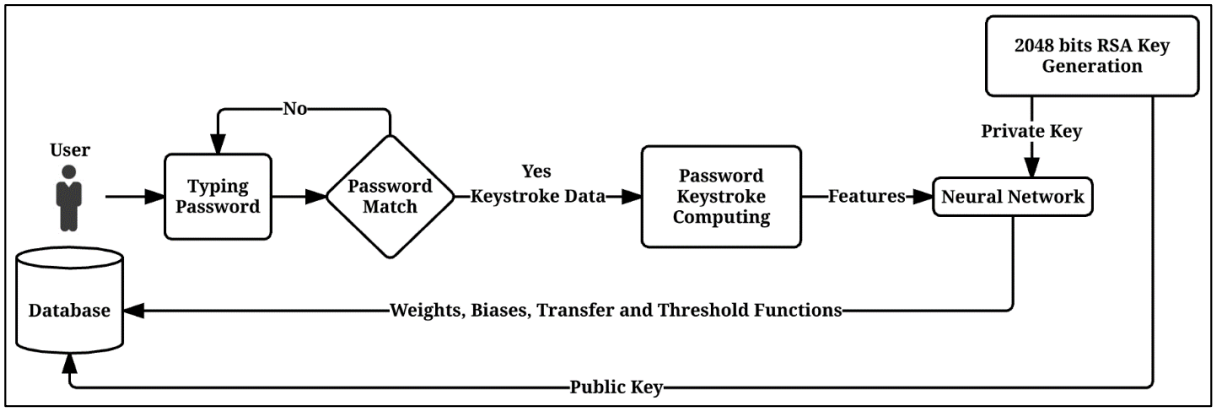


Figure 3. 4 Block Diagram of Key Learning Stage (modified from Chang 2012)

On the key regeneration stage, the authentic users generate a private key as long as they input the valid password. The keystroke features are assembled at the test time and the neural network is reconfigured using the stored helper data (i.e. weights). Accordingly, the test features are trained by the neural network to regenerate/identify random private key of 2048-bit for encryption/decryption. For key verification. As a result, the stored public key will be used to verify the validity of the constructive private key and to ensure that they are a pair. Finally, the genuine user can decrypt or sign a ciphered document using an RSA algorithm. Figure 3.5 illustrates the block diagram of the key generation stage:

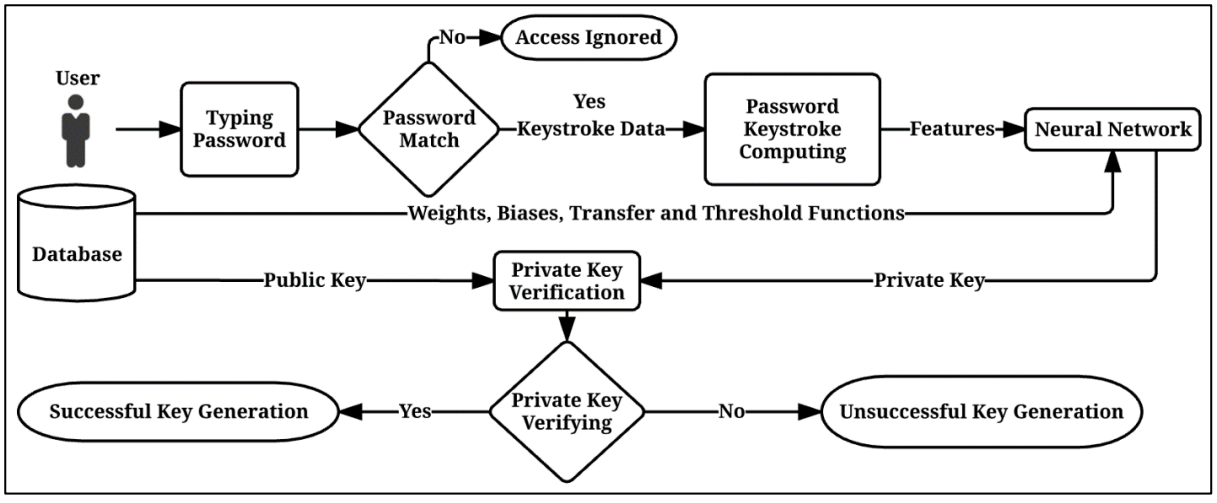


Figure 3. 5 Block Diagrams of Key Generation Stage (modified from Chang 2012)

The experimental work was conducted with 23 respondents as authentic individuals using their strong passwords and 60 hackers to hack the users' accounts. In terms of analysis aspects, the genuine users had 10 attempts to sign into their accounts while the adversaries had 20 trials to attack the valid key; therefore, the analysis sessions were 230 and 27600 respectively. FRR and FAR results were on average 5.25% and 9.61% consecutively, and these are fairly acceptable rates. Nevertheless, there were unacceptable FRR and FAR results for some passwords. For instance, the FAR of the password "630404" was 72%, and the FRR of the password "still531" was 29%. The author claimed that these passwords were weak; therefore, the proposed approach is still unreliable as the features are extracted from typing poor passwords. There was also no evaluation about the biometric entropy; therefore, it is unknown if the presented approach can resist feature guessing attack.

3.4.3 Fuzzy or Secure Sketch Approaches

Overall, fuzzy or secure sketch can take advantage from the fuzzy commitment concepts in using ECC to eradicate the biometric variabilities as well as it biometrically forms secure or fuzzy entity which can be used for security aims.

Conceptually, the elementary fuzzy sketch framework was introduced by Dodis et al. (2004) who claimed that a fuzzy sketch/entity can be extracted from the source biometric template of the genuine user without any leakage of this template. Subsequently, the test biometric template from the same user is presented to regenerate the same fuzzy sketch, and the verification process will be valid if these secure sketches are equal to a specific threshold.

Several researchers have published secure sketch frameworks depending on ECC to treat the biometric differences. Among these, Kanade et al. (2009a) presented a fuzzy sketch framework by using the iris biometric and the combination of RS and Hadamard codes. Of course, RS can adjust the inherent changes in biometrics, such as placing the biometrics on the capture device differently, whilst Hadamard codes remove the errors that are caused by the capture device. In this research, the binary iris codes are extracted using the Open Source Iris Recognition System (OSIRIS). Subsequently, the scrambling approach randomises the source iris code ϕ_{source} by using a password to obtain the modified iris code ϕ'_{source}, aimed at increasing the security aspects. For verification purposes, a random outer key K is hashed by one-way hash function, and simultaneously encoded with the Hadamard and RS codes to produce the pseudo-iris code ϕ_{ps}. In addition, with the aim of raising the rate of error correction over 35%, the procedures of additional zeros insertion and truncation is used to add 792 values in which two zeros are inserted after every three iris bits. Accordingly, the XOR bitwise operation is implemented between the modified iris code ϕ'_{source} and the pseudo-iris code ϕ_{ps} to maximise the security, where the locked iris code ϕ_{lock} is obtained via:

$$\phi_{lock} = \phi_{ps} \oplus \phi'_{source} \quad \ldots\ldots\ldots (4)$$

At end, the AES cryptosystem ciphers the locked iris code ϕ_{lock} and the password of the scrambling approach for further security. These ciphered data as well as the hashed value $H(K)$ are stored on a smartcard as helper data.

On authentication, the same steps as at the time of registration are used to form the modified target iris code ϕ'_{target}, and then the modified pseudo-iris code ϕ'_{ps} is obtained by:

$$\phi'_{ps} = \phi_{lock} \oplus \phi'_{target} \quad \ldots\ldots\ldots (5)$$

Equation (4) can be substituted in (5) to conclude with the following:

$\phi'_{ps} = \phi_{ps} \oplus \phi'_{source} \oplus \phi'_{target}$

$\phi'_{ps} = \phi_{ps} \oplus e$

The e value represents the difference between the iris codes. Of course, ϕ'_{ps} is decoded using Hadamard and RS codes to recover the external key K'. If $H(K)$ equals $H(K')$, then K' will be processed further to reconstruct the source iris code. In the next step, the recovered external key K' is encoded once again by Hadamard and RS codes to obtain the regenerated modified pseudo-iris code ϕ''_{ps}. This ϕ''_{ps} is used to unlock the ϕ_{lock}, and to retrieve the regenerated modified source iris code ϕ''_{source} via:

$\phi''_{source} = \phi_{lock} \oplus \phi''_{ps}$ . . . . . (6)

Since, $H(K) = H(K')$, $K' = K$, and $\phi_{ps} = \phi''_{ps}$ then,

$\phi''_{source} = \phi_{lock} \oplus \phi_{ps} = \phi'_{source}$

As mentioned previously, the ϕ''_{source} is compounded with additive zeros, so the truncation procedure is applied to erase these zeros. Finally, a descrambling approach is applied on ϕ''_{source} by using the stored password to obtain the regenerated reference iris code ϕ_{reg}.

The experimental work of this research was performed on the public database Iris Challenge Evaluations (ICE) that included 244 respondents. The accuracy results were 0.76% FRR and 0.096% FAR and the total entropy was 94 bits where the entropy of the regenerated iris template was approximately 42-bit and the entropy of the password was 52-bit. The researches have realized the difference between the bitlength of the biometric key and its effective bitlength in resisting brute force attacks

where the entropy factor was evaluated reliant upon the fact that not all values had an equal probability.

In the same situation, Sutcu et al. (2007) and Cimato et al. (2009) used ECC in secure sketch-based multibiometric mode to handle the biometric variances. The study of Sutcu et al. (2007) introduced a fuzzy sketch approach using the fingerprint minutiae and the face biometrics. Having established the features appropriately from the minutiae points and the face samples by using the geometrical transform and singular value decomposition methods respectively, normalization and binarization techniques were employed to transform them from the real time into normal distributions. Accordingly, a similarity relation-based error correction code which was experimentally determined between the source and test features was applied to produce a consistent fuzzy/secure construction (Li et al., 2006).

In Sutcu et al.'s contribution, once the binarized feature vectors are established, the combination feature level based on AND bitwise operation is applied to them to obtain the multibiometric template. Subsequently, the secure sketch is constructed via calculating the difference between the error correction codes and the resultant multibiometric template. Respecting the analysis and evaluation, the databases of NIST for the fingerprint and Essex Faces94 for the face are used for empirical works (Garris and McCabe, 2000, Spacek, 2007). The NIST database included 258 respondents, and Essex Faces94 database 152 participants. The researchers evaluated the effective bitlength of the biometric data where the lower entropy for the fuzzy sketch was 39 bits at recognition performance of 2% FAR and 1.4% FRR. The imperfection of this research was the lack of evaluations of within-class variations.

Likewise, Cimato et al. (2009) posed a multibiometric secure sketch system to increase the security and the accuracy purposes by multiple instances of the same biometric, right and left irises. On registration, the right iris template *RI* is encrypted by the AES cryptosystem, and concurrently hashed by using SHA-1 for verification goals in which *H(RI)* is obtained. Subsequently, the output of AES algorithm is encoded by RS codes to obtain the codewords *cw* that will be xored with the left iris codes *LI* to obtain the final template ϕ. At this time both *H(RI)* and ϕ are stored as helper data. On authentication, the refresh left iris template *LI´* is xored with the stored template ϕ to recover the codewords *cw´*. Of course, the codewords of *cw´* and *cw* should be in a tolerable range of error, where their changing should not affect the user's biometric authentication. Consequently, both RS decoding and AES decryption are applied to *cw´*, which represents (RS encoding (AES encryption (RI))), to regenerate the reference right iris template *RI*. Accordingly, a comparison operation is carried out between the hashed value of the constructive and the stored iris templates. If they are equal, then the verification process has been successful.

The suggested system was tested by using the public database CASIA which included 108 participants. The bitlength of the right iris template was 9600 bits. This bitlength accomplished good separability between the genuine user and the attacker where the EER value was 0.5%. In contrast, the bitlength of the left iris template was 1920 bits, and this bitlength performed weakly in separating genuine and malicious users at 9.9% EER. On the whole, the EER value of the proposed multimodal secure sketch system was 0.96%. However, there was no empirical evaluation regarding the biometric entropy.

Apart from the above approaches, some contributions have used specific secure solutions to protect the extracted biometric key, whether it is stored somewhere in a storage unit or transmitted into a third trusted parity (Rathgeb and Busch, 2012). Among these contributions, Mohanavalli et al. (2014) suggested a bio-crypto key generation system for encryption and decryption facilities using the fingerprint biometric. This system dealt with the issues of multiple keys generation and randomisation, and intra-class variations from the same biometric. The authors claimed that if the biometric key is only extracted from the source biometric sample at both parties (i.e. sender and receiver), then the same key will be certainly produced without any differences. Firstly, the biometric features are created via Discrete Wavelet Transformation (DWT). Afterwards, Keccak hash function is applied on these features to generate the biometric key *K* which will be employed to cipher a particular document *D*. Accordingly, the receiver will need *K* to decipher *D*; therefore, the same features which are used to generate the key *K* will be encrypted by RSA algorithm, and transmitted to the other party. Keccak hash function will be applied upon the fingerprint features for multiple key generation and randomization to resolve the problem of key hacking. Of course, RSA will require a pair of keys; therefore, One-Time Password (OTP) is used as an input to the key generation process of RSA to generate these keys. However, this research does not demonstrate the key length and its entropy to illustrate the system feasibility in terms of security.

Several works have been published to present conceptual and analytical understanding of the art of biometric encryption – particularly of biometric key generation. Merkle et al. (2012) presented a comprehensive approach to analyse

multimodal biometric fusions, which are feature level, score level and decision level fusions, in terms of their effects on recognition performance and security in biometric encryption systems. In addition, the authors proposed hash-level fusion as an efficient direction to combine the biometric features conceptually. Merkle et al. (2012) claimed that the multibiometrics consolidation of the bio-cryptographic approaches, such as the key generation could be applied depending on the hash level because the unimodal of bio-cryptosystems relies upon the successful verification of equality hash value. Hypothetically, the hash function is implemented individually upon the extracted features from the both biometric modals during the registration phase. As a result, two independent parts of public data P_1 and P_2 and two bit strings b_1 and b_2 are produced where b_1 and b_2 are merged before the hashing process. On authentication, the stored hash value of $H(b_1/b_2)$ is verified with the generated bit strings bb_1 and bb_2 from the target biometric modals.

Another analytical work was introduced by Golic and Baltatu (2008) who claimed that the statistical independence of random variables cannot be gauged via the average min-entropy metric. They used as an alternative for measuring the entropy of the biometric key the conditional Shannon entropy criteria. Of course, conditional Shannon entropy and the average min-entropy can estimate the security of biometric cryptosystem from the viewpoint of information theory. Nonetheless, these assessments do not show the actual values of biometric data when drawing their probabilities only. As a result, these metrics should not be used only to analyse the security of biometric encryption systems as these systems could not be secure from the information theory aspects, but are mathematically secure. Alongside this work, Jain et al. (2008) introduced a conceptual review upon the schemes of biometric

template protection specifically on biometric key generation. The writers essentially elaborated the biometric key generation concepts in terms of advantages, disadvantages, open challenges, and main approaches. Further, the authors claimed that the lack of high discriminable biometric features will influence the performance of the key generation system with regard to key stability and key entropy. In particular, these concepts are defined by Jain et al. (2008) as:

> "*Key stability refers to the extent to which the key generated from the biometric data is repeatable. Key entropy relates to the number of possible keys that can be generated.*"

An analytical and assessable work on the biometric key generation approaches is illustrated in Table 3.3 as follows:

Table 3. 3 Analytical and Assessable Work on Biometric Key Generation Approaches-Based Timeline

#	Authors & Year	Biometrics	Fusion	Feature Extraction	Generation	Verification	Performance	Users	Variances Treatment	Type
1	Bodo 1994	—	—	—	—	—	—	—	—	Conceptual
2	Juels and Wattenberg 1999	Fingerprint	—	—	Commitment Function	Hash Function	—	—	ECC	Conceptual
3	Janbandhu and Siyal 2001	Iris	—	Iriscan Product	RSA and DSA	Decryption Process	Key RSA= 512-Byte Key DSA = 160-Bit EER= 1in 1.2 Million	—	—	Prototype
4	Dodis et al. 2004	Fingerprint	—	—	Biometric Combination	Hamming Difference	—	—	ECC	Conceptual

5	Hoque et al. 2005	Handwritten Signature	—	—	Vector Quantization	Standard Deviation	Key= 32-Bit FAR=35.2% FRR=5.6%	144	—	Simulation
6	Teoh and Kim 2007	Fingerprint	—	Gabor Transform	Randomized Dynamic Quantization	Hamming Difference	Key=375-Bit FAR=0% FRR=0.9%	100	RS	Real
7	Chang and Roy 2007	Fingerprint	—	PCA	Biometric Combination	Maximum Likelihood	Key=8-10 Bits FNMR=0.09%	4000	ECC	Real
8	Sutcu et al. 2007	Fingerprint and Face	And	Geometric Transform and SVD	Continuous Secure Sketch	Hash Function	Key Entropy= 39-Bit FAR=2.000009% FRR=1.4%	258	ECC	Real
9	Sheng et al. 2008	Handwritten Signature	—	Statistical Approach	Genetic Algorithm	Logarithm Formula	Key=20-Bit FAR=0% FRR=14.5%	359	—	Real
10	Kanade et al. 2008	Iris	—	OSIRIS	Randomized ECC-Based xor	Hamming Difference	Key=128-Bit FAR=0.55% FRR=1.04%	124	RS & H	Real
11	Sutcu et al. 2008	Fingerprint	—	PCA	Probabilistic Scheme of User-Specific Cuboid	Hash Function	Key=30-Bit FAR=1.0019% FRR=0.11%	1035	LDPC	Real
12	Golic and Baltatu 2008	—	—	—	—	—	—	—	—	Analytical
13	Jain et al. 2008	Fingerprint-Iris-Face-Palmprint-Handwritten Signature	—	—	—	—	—	—	—	Review
14	Atah and Howells 2009	Voice	—	Microsoft and Matlab Tools	Quantization-Based Binary System Conversion	Standard Deviation	Key=32-Bit	106	Normalisation	Real
15	Kanade et al. 2009	Right and Left Irises	Xor	OSIRIS	Randomized ECC-Based xor	Hamming Difference	Key=147-Bit FAR=0% FRR=0.18%	120	RS & H	Real
16	Kanade et al. 2009	Iris	—	OSIRIS	Randomized ECC-Based xor and AES	Hash Function	Key= 186-Bit Key Entropy= 94-Bit FAR= 0.096% FRR= 0.76	244	RS & H	Real
17	Cimato et al. 2009	Right and Left Irises	Xor	Open Source Code	ECC-Based AES and xor	SHA-1	Right=9600-Bit Left=1920-Bit EER=0.96%	108	RS	Real
18	Jagadeesan and Duraiswamy 2010	Fingerprint and Iris	Concatenation	Morphological Transform and Daugman Iris Technology	Scrambling Method-Based Mod 2 Operation	Hamming Difference	Key=256-Bit	—	Preprocessing	Real

19	Balakumar and Venkatesan 2011	Fingerprint and Iris	Concatenation	Morphological Transform and Daugman Iris Technology	Scrambling Method-Based Mod 2	Hamming Difference	Key=256-Bit FAR=0.2351% FRR=85.07%	100	Preprocessing	Real
20	Li et al. 2012	Fingerprint	—	PCA	Enhanced User-Specific Cuboid	Hash Function	Key=48-Bit FAR=0% FRR=4.85%	110	LDPC	Real
21	Feng and Yuen 2012	Face	—	Fisherface	Binary Discriminant Analysis	Euclidean Difference	Key Entropy= 75-Bit Recognition Rate= 96.38% Time=178.3 Seconds	68	BCH	Real
22	Chang 2012	Keystroke	—	Password Keystroke Computing	Neural Network	Private Key Generation	Key=2048-Bits FAR=9.61% FRR=5.25%	23	—	Real
23	Merkle at el. 2012	—	Hash Fusion	—	—	—	—	—	—	Analytical
24	Wahdan et al. 2013	Iris, Face and Voice	AES	DWT, DCT and Cepstrum Analysis	ECC and Binarisation-Based Threshold	Document Decryption	Key=192-Bit Accuracy =99.83%	20	RS	Real
25	Chin et al. 2014	Fingerprint and Palmprint	Rearrangement	Gabor Transform	Random Tiling Transform-Based Binarisation	Hamming Difference	Features= 200 Dimensions EER=0.05%	300	Preprocessing	Real
26	Mohanavalli et al. 2014	Fingerprint	—	DWT	Kaccak Hash Function	IDWT	—	—	—	Prototype
27	Abuguba et al. 2015	Iris and Face	Substitution	Daugman Iris Technology and PCA	Normalization -Based Substitution	Hamming Difference	Key=256-Bit FRR=12.51%	40	Preprocessing	Real

On the whole, a significant research has been presented within the topic of biometric key generation. Although the direct key generation schemes are still open to challenge, the approaches of indirect key generation can introduce encouraging directions for novel developments. It is clear that the approaches of indirect key generation have concentrated upon the most reliable biometrics, such as fingerprint, and iris aimed at rising up the accuracy aspect. Furthermore, some approaches have

exploited off-the-shelf biometric products to extract the robust biometric features, and to generate the consistent biometric keys over time. Also, the key generation methods are applied depending on statistical approaches, Artificial Intelligence (AI) approaches, and randomisation approaches. From the perspective of verification, specific processes, such as hash functions, distance metrics and successful decryption are utilised to verify the valid biometric key. Particular error correction codes are employed to reduce the intra-person variances, such as RS, BCH, and LDPC, and the latter outperforms the others in maximising the performance of the proposed systems. It is noteworthy that the performance results with regard to FAR, FRR and ERR on average were fairly acceptable.

3.5 Biometric Key Binding Approaches

Significant amounts of research have jointly incorporated biometric and cryptographic principles to develop secure and complementary approaches. One of these is biometric key binding which binds an outer cryptographic key with the source biometric template to produce helper or public data during the registration stage (Jain et al., 2008). On verification, the reference biometric template and the secret key are unbound to employ the latter for cryptographic goals if the test and the source biometric templates are from the same genuine user (Rathgeb and Uhl, 2011). Of course, the computational complexity of the integration process does not reveal important information about the biometric template and the cryptographic key. Overall, biometric key binding poses some good attributes. One of these aspects is that the same cryptographic key should be regenerated at the verification phase because it is not derived from the biometric templates. Further, the use of an outer secret key leads to low FAR results, and certainly will achieve the revocability

requirement of the biometric cryptography where it is easy to revoke the compromised biometric template and reissue different one by choosing another secret key. On the other hand, there is an aspect which can be considered a challenge. This aspect is that the helper data should be generated cautiously to maximise the security and accuracy requirements. Once the attacker has analysed the helper data and guessed the key and the biometric data, the biometric key binding system will be useless (Uludag et al., 2004, Jain et al., 2008).

The approaches of biometric key binding can be classified into conventional approaches, fuzzy commitment, fuzzy vault, and salting or BioHashing approaches. These approaches are discussed in the following subsections.

3.5.1 Conventional Approaches

The traditional approaches were the first attempts by biometric encryption scientists to comprehend the research area of biometric key binding. The first approach to biometric key binding was suggested by Tomko et al. (1996) who proposed a public key cryptographic system depending on the fingerprint biometrics. In a brief review, on registration, certain fingerprint signals are initially selected to compose a distinctive value which is exploited to construct random numbers by using a random number generator. Simultaneously, the biometric features are taken out from the fingerprint singles via Fourier Transform (FT). Thereafter, fingerprint features are bound with the random numbers by filtering approach. Accordingly, this integration will be stored on the smartcard of the authentic user. On the probe stage, the biometric features are also extracted from the query fingerprint sample by using FT. After feature extraction, a correlation mechanism is applied in order to compare

between the source and the query features. The random numbers will be valid to be utilised in the key generation phase of a public cryptosystem depending on the comparison operation.

In a similar way, Soutar et al. (1998) applied a filter generator based on Discrete Fourier Transform (DFT) to extract biometric features from subsequent fingerprint acquisitions of the same user (i.e. 6 training samples). A mathematical formula is accordingly designed by utilising the correlation concepts to find the peak of distinctive features among them. The authors claimed that this procedure will cope with the discrimination and distortion of the fingerprint features. For further reducing the biometric variances, the researchers selected the core 64×64 of the features and converted them into binary values using a binarization-based threshold method. The binary features are then incorporated with an external secret key to constitute a secure entity using specific permutation approach. For key retrieval, a lookup table is constructed and stored in order to recover the same cryptographic key reliant upon the test fingerprint features. For verification purposes, the key ciphers N bits of the binary features using 3DES algorithm, and these encrypted N bits are hashed using a SHA-1 algorithm to construct a validation code vc1 which is also stored at some location. On authentication, the cryptographic key is extracted to generate another validation code vc2 to be compared with vc1. If vc2 $\neq$ vc1 then the cryptographic key is invalid; otherwise, it will be used for encrypting/decrypting data.

Generally, Tomko et al.'s and Soutar et al.'s contributions pose a number of imperfections. One of these imperfections is that the contributors presumed that the fingerprints database was entirely rectified. Image rectification or image alignment is a transformation process to project two or more different images, one termed source

image and the others are termed the target images, into a common image version (Chung et al., 2005). In a realistic environment, the advanced sensors of the capture devices can contribute to acquire images of high quality; however, these high quality images contain on a number of negative attributes (Crisp, 2013). Consequently, the captured samples must be aligned to eliminate the emerging irregularities. Another imperfection is that the researchers did not illustrate a strict security analysis to provide conclusive evidence about the strength of the security and to convince the beneficiaries to adopt the proposed systems. What is more, there were no verification results to demonstrate the accuracy of the suggested schemes in terms of FAR and FRR to motivate stakeholders to distribute these technologies in the industrial world.

3.5.2 Fuzzy Commitment Approaches

Generally, fuzzy commitment uses the error correction codes to handle the variations of biometric. In addition, biometric key binding-based fuzzy commitment integrates an external key with the error correction codes and the biometric template to generate secure entity. Fuzzy commitment approaches need usable, accurate and fast recognition performance-based constant biometrics to tolerate as much as possible the differences between the source and the test biometric templates. The iris biometric seems the most appropriate modality to attain these characteristics (Janbandhu and Siyal, 2001).

Particular fuzzy commitment schemas are proposed via iris, and among these are the approach that was introduced for smartcards by Hao et al. (2006). In this contribution, a 256-byte source iris template is formed by using 2D Gabor wavelet

transform, and then a 140-bit outer secret key is encoded with concatenated Hadamard and Reed-Solomon codes to obtain symmetric 256-byte codes for the iris codes. Accordingly, the integration process is applied between these codes via the xor bitwise operation to obtain the lock codes. During the authentication phase, the same steps as those described above are applied to the test iris sample. The cryptographic key will be recovered if the difference between the source and the live biometric templates is lesser than or equal to the Hamming distance divided by two. The suggested system was tested using a proprietary database that included 700 iris samples of 70 participants, with 10 samples from each iris where all the iris samples were captured in standard settings via the same CDD camera at a fixed measurement distance. The authors demonstrated positive FAR and FRR results of 0% and 0.47% respectively with a retrieval cryptographic key of 140 bits.

In a follow-up approach, Sukarno et al. (2009) presented a fuzzy commitment system using the incorporation of Reed-Solomon and Hadamard error correction codes to maximise the Email security in mobile devices. The researchers constructed a 9600-bit reference iris template $\emptyset_{Ref}$ via a Libor Masek algorithm. On the other hand, an external secret key K is randomly generated and hashed by the SHA-512 algorithm for verification purposes, and concurrently it is encoded with RS and Hadamard codes to obtain the pseudo iris template $\emptyset_{PS}$. Accordingly, the XOR bitwise operation is applied between $\emptyset_{Ref}$ and $\emptyset_{PS}$ to produce the locked template $\emptyset_L$. Further, an AES cryptographic algorithm is used to encrypt the locked template $E(\emptyset_L)$ in order to maximise the privacy. On verification, AES cryptographic algorithm decrypts the stored $E(\emptyset_L)$. At the same time, the query iris template $\emptyset_{Query}$ is presented to unlock $\emptyset_L$ using XOR bitwise, and this will contribute to obtain the pseudo iris template $\emptyset'_{PS}$

once again. As such, $Ø'_{PS}$ is decoded by RS and Hadamard codes to retrieve the outer key K'. The key verification will be successful if the stored hashed value equals to the hashed value of the retrieved key $H(K')$. The experiments of this research were carried out on proprietary database which included 70 participants with 10 samples from each one. The bitlength of the retrieval key was 408 bits and the FAR and FRR results were 0% and 1.5873% respectively. However, the key of AES will be stored somewhere; therefore, the security of this key will also depend upon traditional passwords.

Interestingly, Ziauddin and Dailey (2010) suggested fuzzy commitment schema by utilizing the Bose–Chaudhuri–Hocquenghem (BCH) error correction codes. At the registration phase, three independent iris samples are captured to create three basis templates of 9600 bits, based on a Masek and Kovesi algorithm. Accordingly, with the purpose of improving the recognition performance and eliminating the iris variations, the authors empirically investigated the corrupted bits and the inconsistent positions within the basis templates. As a consequence, these inconsistencies are masked out via masking manners-based Hamming distance metric. Moreover, a 9600-bit indicator vector I is constructed to indicate into the locations of the most consistent bits amongst the basis templates, and then this indicator is stored to figure out the reliable bit locations upon the target iris template. Subsequently, 4095-bit uniform final template UFT is obtained from the stable bits which are not corrupted and identical over all the basis templates. Besides this, in order to further reduce the biometric variances, BCH error correction codes are encoded with a 260-bit external key K, and then the resultant codes O are xored with the UFT to originate the retrieval information RI that contributes to recover that outer

key. At the same time, the cryptographic key of 260 bits is hashed for verification purposes. At end, the indicator vector *I*, the hashed key *H(K)*, and the retrieval information *RI* are stored on the smart card as helper data to contribute in verifying the genuine user. The experiments of this research were conducted using the iris database of Bath University. The free version of this database included 1000 iris images of high quality which are captured from 25 participants.

Ziauddin and Dailey's contribution introduces interesting aspects. One of these aspects is that the researchers enhanced the Masek and Kovesi algorithm of iris template creation by using image blurring to reduce the intra-person variations. Image blurring reduces the edge details and regulates the differences in curves and lines (Gonzalez and Woods, 2008). This perspective will minimize the capacity of the error correction codes that should be used to handle the biometric differences; therefore, the bitlength of the outer cryptographic key can be increased, and the bitlength of the recovered key which is 260 bits confirms this perspective. Another good aspect is that the combination of the consistent and masked bits presents the best accuracy results where the FRR and FAR rates were 0. What is more, the contributors avoid the issue of rotational inconsistencies which take place due to the acquisition of images by a rotation process. However, the stored hashed key on the smart card may be considered the security issue when the imposter has the ability to break the used hash function. Figure 3.6 shows the block diagram of the enrolment and verification processes:

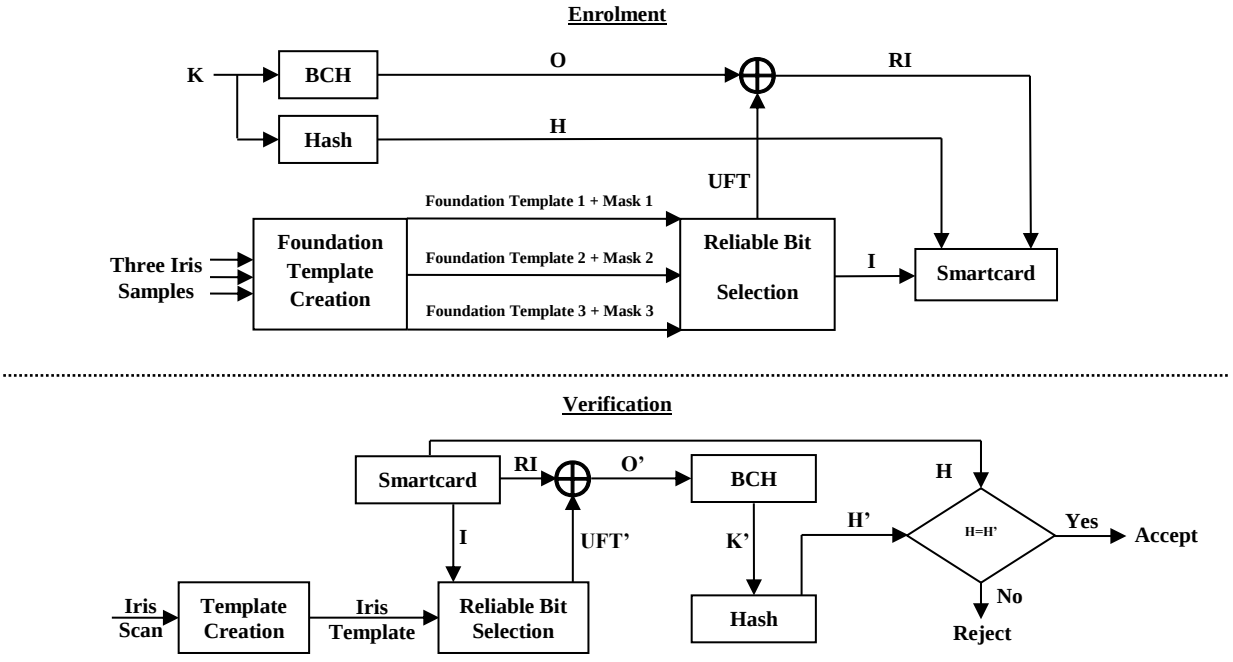


Figure 3. 6 Block Diagram of Enrolment and Verification (modified from Ziauddin and Dailey 2010)

3.5.3 Fuzzy Vault Approaches

The fuzzy vault approach is the same as in fuzzy commitment, where it utilises the error correction codes to treat the biometric variabilities. Fuzzy vault, however, secures the biometric template and the cryptographic key mathematically against forgers. The primary approach was suggested by Juels and Sudan (2002) who stated that biometric can be bound with an outer key to generate a vault based on mathematical polynomial construction, where neither the biometric nor the key can be guessed easily by imposters. A set of biometric features $x = \{x_0, \dots, x_{s-1}\}$ is created with error correction codes, and then an external secret key of $k_0 \dots k_{m-1}$ is chosen to calculate a polynomial of $p(x) = k_{m-1}x^{m-1} + \dots + k_1x + k_0$ at each element in x to generate a set of authentic points $\{x_i, p(x_i)\}$ where $m-1 \leqslant s-1$. In addition, a set of chaff points is inserted onto a polynomial in order to camouflage the authentic points, and afterward both chaff and authentic points compose the fuzzy vault. During the

authentication phase, the live biometric template should considerably overlap with the source template to reconstruct the polynomial structure via error correction codes and to recover the biometric template and the cryptographic key.

On the second contribution in fuzzy vault, Clancy et al. (2003) presented the first clear fuzzy vault system practically by using the features of fingerprint minutiae where the authors statistically investigated the optimum parameters for fuzzy vault construction. At first, a pre-processing stage is applied to align the fingerprint samples into a normal distribution and to eradicate the biometric noises. The minutiae points are projected onto the polynomial construction to be integrated with 128-bit outer secret key, and accordingly a set of chaff points are inserted at random to create the fuzzy vault paradigm depending on RS codes. Subsequently, on verification, the target fingerprint template is presented, where it should overlap with the reference template to reconstruct the polynomial structure and recover the secure elements, the biometric template and a 128-bit cryptographic key. The proposed approach is performed on a real fingerprint dataset to demonstrate the feasibility of this approach. The authors reported that the security of the primary fuzzy vault counstruction which is presented by Juels and Sudan (2002) cannot be visualized with reality. The recognition performance of FAR and FRR was 0% and around 25% respectively. However, this accuracy figure reflects that the suggested system may not recognise the quarter of the legitimate users.

In more efforts, with the aim of dealing with the biometric variations prior to aligning each biometric sample, Nandakumar et al. (2007) suggested a fuzzy vault schema in which a post processing step is performed at the time of verification. The research claimed that because the fuzzy vault system stores the transformed version of the

source biometric template only, the alignment process with the transformed version poses a significant challenge. Therefore, a set of reliable curvature points of the fingerprint biometric is generated from their orientation and employed as helper data to align the reference and the test biometric templates, based on a statistical approach and geometrical transformation. Commonly, the appendage process between the orientations of curvature points and a 128-bit external key is applied by the mathematical polynomial structure with the assistance of the error correction codes and chaff points. The experiments for this research were conducted on a fingerprint database available publicly from FVC2002 website. The proposed system demonstrated promising results of performance, where FAR and FRR values were 0.04% and 4% respectively with a retrievable key of 128-bit.

In further attempts, Khalil-Hani et al. (2013) illustrated that although the meaninglessly creation of chaff points are used to camouflage the outer cryptographic key on the biometric template effectively, it is deemed the most complex process of fuzzy vault construction. The researchers enhanced the complex algorithm of Clancy et al. (2003) by using the fingerprint biometric. In this approach, the external cryptographic key is encoded by cyclic redundancy check codes for verification purposes, and then secured via simple randomisation method-based XOR bitwise. After the randomisation process, this key is divided into small variables of size n+1, and these variables V became the polynomial coefficients of a degree n: $P(X) = V_1X^n + V_2X^{n-1} + \ldots + V_nX^{n+1}$. Accordingly, fingerprint features of 16 bits are mapped onto the polynomial as x coordinate values to be used to calculate the y-coordinate values. The x and y coordinate values are considered the valid points of the fuzzy vault. From the other perspective, the algorithm of Clancy et al. (2003) is

enhanced via specific ideas with the purpose of speeding up the chaff points generation.

Amongst these ideas, the circle packing theorem is used to geometrically pack a number of circles in a surface according to a number of conditions. These conditions are that the circles of the same radius should not be overlapped and should be packed into a square, circle or triangular figure. That is, the new points are inserted into the vault template if their boundaries do not overlap the boundaries of the other present points. Another insight is that the subtraction, addition, and comparison operations are used only to simplify the process of chaff generation rather than the square root operators that add a computational complexity to the system. Subsequently, the query fingerprint template is compared with the members of the fuzzy vault as x-coordinates to obtain the closest vault members by using the steps of the integration process, where the Gaussian transform is used to recreate the polynomial structure.

The empirical results via the circles with the boundaries of smooth edges reported negative results concerning the overlapping boundaries; therefore, the developers employed the squares to cope with the overlapping issue. There was, nevertheless, a need to test the boundaries of the new points to ascertain whether or not they overlap the other present points over time. In addition, the executed time of the presented chaff generation was 310 seconds which is faster than the execution time of the Clancy algorithm at chaff and minutiae points of 500 and 30 respectively. What is more, the bitlength of the outer cryptographic key was 128 bits which are appropriate for AES encryption. However, this research did not provide a strict security analysis as a conclusive evidence about the strength of the security so as

to convince potential beneficiaries to adopt the proposed system. There were also no results to demonstrate the accuracy of the suggested schemes in terms of FAR and FRR to motivate stakeholders to introduce these technologies to the industrial world.

So as to cope with the imperfections in the works of Clancy et al. (2003) and Khalil-Hani et al. (2013), Nguyen et al., (2013) introduced a quick chaff generation algorithm based on the ideas of Clancy et al. (2003). In the proposed algorithm, the fingerprint sample is separated into a number of cells (pixels) where each cell is located beside eight adjacent cells. Subsequently, a fresh chaff point is created at random according to a number of conditions. The first condition is that the unique chaff point for a given image cell is created at random; however, if the image cell includes a chaff point or genuine point then this cell should be ignored. Secondly, the Euclidean distance between the fresh chaff point and the existing eight points is greater than or equal to the given distance threshold. After the generation of chaff points, an image cell matrix is employed to investigate whether or not the correct points and chaff points exist. The experimental work was carried out on the public databases from FVC2002 website, which included 100 respondents with eight samples for each respondent. The experiments confirmed that the average EER for database of FVC2002-DB1A was 2.4%, while the EER result for the other one of FVC2002-DB2A was 1.9% on average. Further, the authors succeeded to retrieve a 128-bit external cryptographic key from the vault construction. When creating 24 fingerprint minutiae points and 240 chaff points, the Nguyen's algorithm was quicker than the algorithms of Khalil-Hani et al. (2013) and Clancy et al. (2003) with 41.86 and 14.82 times respectively.

In more recent research, Li et al. (2015) presented an analytical paradigm to combine the computational complexity along with the entropy for comprehensive security analysis. Moreover, the contributors proposed a fuzzy vault system via multiple instances of the fingerprint biometric. Overall, the idea of this system is to add computational complexity to thwart attackers who attempt to hack the biometric key. A Delaunay triangulation transform based on Voronoi diagram is used to extract the fingerprint minutiae set. In this approach, the fingerprint image is divided into a number of small triangular regions via a Voronoi diagram, where the entire points of a region are placed at the close minutia. Subsequently, the minutiae in adjacent regions for the whole Voronoi diagram are connected to form the Delaunay triangulation net. Accordingly, the feature vector is extracted by using a number of geometrical transformations. With regards to the fusion manner, the templates from different fingerprints are securely fused through two levels. During the first level, each template is hashed individually by hash function, and then the fuzzy vault scheme-based polynomial construction is applied to bind each hashed template to a user-specific key that must be constant for all templates because it will be used in the next level of fusion. Also, the hash function is applied on these keys and then all the templates are jointly hashed for verification purposes. At the second level, an additional security control is added into the user-specific keys by the means of Shamir's secret sharing approach, where each key is divided into a number of shares and combined based on the mathematical polynomial construction. In terms of analysis and evaluation, the fingerprint images were collected personally to construct a database of 150 participants. The researchers achieved a biometric key of 32-bit entropy with high computational complexity, which makes the system more secure.

Furthermore, the proposed fuzzy vault system demonstrated good recognition performance of 2.67% FRR and 0% FAR.

3.5.4 Salting or BioHashing Approaches

The main objective of the salting or BioHashing approach is to construct an irreversible biometric version with the purpose of protecting the source biometric template. In these approaches, Personal Identification Numbers (PINs) should be also introduced into feature extraction methods at the time of unbinding to construct biometric hashes (Rathgeb and Uhl 2011). In accordance with this, Song et al. (2008) proposed a biohashing approach applicable to smartcards using the biometric hashing method, i.e., one-way feature transformation. In this approach, the biometric hashing method extracts the fingerprint features as a vector by filtering method. Furthermore, a tokenized PIN is used to generate a random vector with the aim of accomplishing the requirements of revocability and key diversity. Subsequently, the inner product is applied between the feature vector of the valid user and the tokenized random number vector, and the result is binarized iteratively using particular criteria to generate irreversible binary features. Following this, Reed-Solomon error correction codes are applied to correct the error between the source and the target fingerprint templates on verification. After RS coding, a 180-bit outer cryptographic key K_0 is xored with the binary features to generate an irreversible template version called “Biocode”. During the validation process, the methods of both biometric hashing and Reed-Solomon are applied on the target fingerprint sample to construct a one-way target template which is xored with the “Biocode” to extract the key K_1. Finally, the key recovery process is accomplished by comparing between the hash function the outer cryptographic key K_0 and the extracted key K_1.

The empirical tests were carried out using three databases from the FVC2002 website, and all of them include eight different positions of one hundred various fingers. Also, the central area 128×128 of each image is determined by pre-processing to diminish the potential variations. On the whole, the research of Song et al.'s may be deemed one of the most interesting papers due to particular characteristics. The authors demonstrated good accuracy results of FAR and FRR which were 0% and 0.827% respectively for all databases. In addition, the source fingerprint template was not stored on the smart card of the authentic user, and this significantly reinforces the security requirements.

In other work, Inthavisas and Sungprasert (2013) presented a salting scheme by using the speech modality in which undesirable features from the frequency-domain are rejected to build an irreversible template; therefore, the forgers cannot recover the original template in case of compromise. The frequency-domain features are extracted by Discrete Fourier Transform (DFT), and one of the training utterances is stored as a keying signal. Subsequently, the Dynamic Time Warping (DTW) method is applied on the rest of the training utterances to produce the robust features where a mapping method-based cepstrum analysis is implemented on these robust features to create Descriptors *D*. Cepstrum analysis is used in mapping method to split up the speech components, which are the excitation source and the vocal tract system components, with the aim of analysing them individually. In addition, the mechanism of random thresholds generation is used in the mapping method to increase the entropy of the speech template in which a number of threshold values, *T* and *TH*, are determined empirically to meet operational conditions. As a result, a number of frequency-domain features are refused or accepted to configure an

irreversible template depending on these operational conditions. That is to say, the Descriptors *D* are constituted if *D* is lower than or equal to *T*. Also, the irreversible template will be constructed if the Distinguishing Descriptors *DD* are lower than or equal *TH*. *DD* are determined by applying the algorithm of sequential backward search on *D*. Accordingly, the smallest variances are selected from *DD* to generate the binary string *S*, and then an initial key *K*, which is picked up by the user, is encoded by BCH error correction codes to produce encoded key *E(K)*. At end, the *S* and *E(K)* are encrypted by the xor bitwise operation to obtain the encrypted data *ED*. Of course, *ED*, the random thresholds, the hash function of the initial key *H(K)* and the irreversible template will be considered as helper data and stored in the storage unit. The helper data will be used during the authentication stage, which will be successful if *H(K)* equals the hash function of the recovered key *H(K')*. Figure 3.7 shows the main processes of the proposed system:

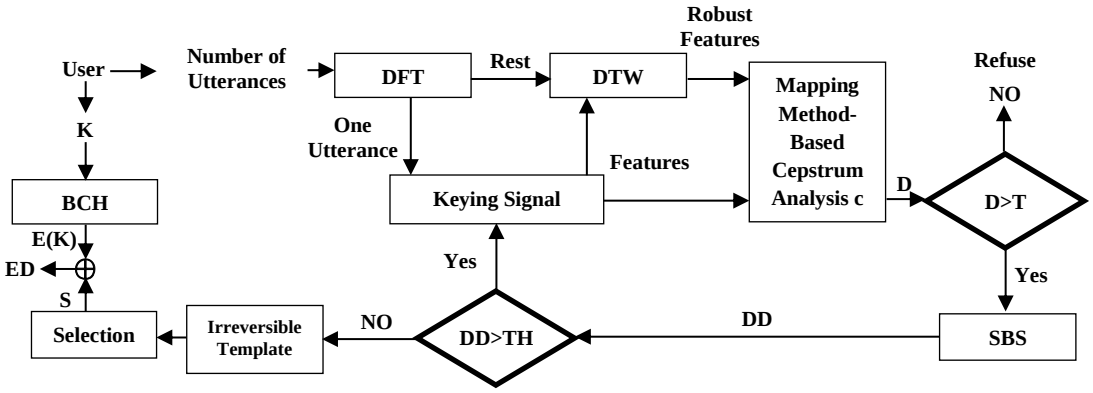


Figure 3. 7 The Main Processes of the Proposed System (modified from Inthavisas and Sungprasert 2013)

48 volunteers participated in the experiments of this research using the public database of the Massachusetts Institute of Technology (MIT). The researchers manifested good performance results where the generated biometric key was 127 bits, and the equal error rates for random legitimate users and forgers were 4.43%

and 13.14% respectively. The only criticism of this contribution is that only seven adversaries were engaged in the experiments.

A chronological and comparative compilation on biometric key binding schemes is illustrated in Table 3.4 as below:

Table 3. 4 Chronological and Comparative Compilation of Biometric Key Binding Schemes

#	Authors & Year	Biometrics	Fusion	Feature Extraction	Integration	Verification	Performance	Users	Variances Treatment	Type
1	Tomko et al. 1996	Fingerprint	—	Fourier Optics	Filter Generator	Optical Correlation	—	—	Physical Optics	Prototype
2	Soutar et al. 1998	Fingerprint	—	Fourier Transform	Filter Generator-Based Binarisation	Decryption Process	Key=128-Bit	—	Preprocessing & ECC	Prototype
3	Juels and Sudan 2002	—	—	—	Polynomial Function-Based Chaff Points	Hash Function	—	—	ECC	Conceptual
4	Clancy et al. 2003	Fingerprint	—	Statistical Approach	Polynomial Function-Based Chaff Points and RS	Hamming Difference	Key=128-Bit FAR=0% FRR=25%	—	RS	Simulation
5	Hao et al. 2006	Iris	—	2D Gabor Wavelet Transform	xor	Hamming Difference	Key= 140-Bit FAR=0% FRR=0.47%	70	RS & H	Real
6	Nandakumar et al. 2007	Fingerprint	—	Statistical Approach and Geometric Transform	Randomized ECC-Based xor	CRC	Key=128-Bit FAR=0.4% FRR=4%	100	Preprocessing	Real
7	Song et al. 2008	Fingerprint	—	Gabor Transform	Biometric Hashing Method-Based xor	Hash Function	Key=180-Bit FAR=0% FRR=0.827 %	100	Preprocessing & RS	Real
8	Sukarno et al. 2009	Iris	—	Libor Masek Algorithm	xor	Hash Function	Key=408-Bit FAR=0% FRR=1.5873%	70	RS & H	Real

9	Ziauddin and Dailey 2010	Iris	—	Improving of Masek and Kovesi Algorithm	Masking Manners-Based ECC and xor	Hash Function	Key=260-Bit FAR=0% FRR=0%	25	Preprocessing & BCH	Real
10	Li et al. 2012	Multiple Fingerprints	Hash Fusion	Delaunay Transform	Polynomial Function-Based Shamir's Secret Sharing	Hash Function	Key Entropy =32-Bit FAR=0% FRR=2.67%	150	—	Real
11	Khalil-Hani et al. 2013	Fingerprint and Vein	Concatenation	Image Processing	Polynomial Function-Based Enhanced Chaff Generation	CRC	Key=128-Bit Time=310 Seconds	—	—	Prototype
12	Nguyen et al. 2013	Fingerprint	—	—	Polynomial Function-Based Fast Chaff Generation	CRC	Key=128-Bit EER for DB1A=2.4% EER for DB2A=1.9% Time=0.8219 Seconds	100	—	Real
13	Inthavisas and Sungprasert 2013	Speech	—	Cepstrum Analysis and DFT	Random Thresholds Generation-Based xor and Sequential Backward Search (SBS)	Hash Function	Key=147-Bit EER =4.43%	48	BCH	Real

As illustrated in table 3.4, the fingerprint and the iris modalities are widely exploited in applying key binding approaches because they are quite consistent throughout. As such, this would escalate the performance of retrieving the bound bio-crypto key from the fingerprint and iris features on the basis of successful biometric verification. Various feature extraction methods are utilised including digital signal processing approaches and statistical approaches. With regard to the binding process, the mathematical polynomial construction is mainly employed to integrate the biometric features with the outer cryptographic key. In addition to this, the one-way hash function mostly verifies the cryptographic key. Interestingly, some researchers incorporated the error correction codes and the pre-processing methods to diminish the intra-person variabilities. Overall, the space of the external secret key ranges

from 125-bit to 408-bit. The experiments of the proposed approaches were conducted on real data collection. Although there are technical research efforts, there is a lack in applying the multibiometric key binding approaches. Furthermore, there is a need to explore the role of the other biometric modalities, such as the face and the keystrokes within the biometric-based key binding mode. However, this needs to investigate fairly constant features over time of face and keystrokes modalities in order to maintain and support the secret key recovery process from those biometric features. Otherwise, the key will not be recovered as long as the constancy of the face and the keystrokes features have been significantly varied. Such biometric key binding techniques built upon different biometric modalities (i.e. fingerprint, face, iris and keystrokes) could possibly introduce positive indications towards developing mature and comprehensive multibiometric-based key binding approaches. Accordingly, these systems would be more robust and accurate in terms of resisting attacks and recovering a constant encryption key on a timely basis. That is, incorporating enormous biometric features from different biometric modalities would reinforce the biometric entropy – thus providing the potential of combating brute force attacks. In addition, if a biometric will be hindered in unbinding the secret key the other one can probably aid in doing so.

3.6 Discussion

With the aim of strengthening the security controls within cloud-based storage in a usable fashion, bio-cryptographic approaches are studied and analysed to develop and investigate a novel encryption approach using transparent biometrics. According to the literature, biometric cryptography can overcome the weaknesses of password verification in more secure and usable way. It offers strong approaches that can be

secure vis-à-vis forgers and the genuine users no longer need to remember long and difficult cryptographic credentials. Of course, a backup mechanism for providing a secure countermeasure should be considered and set in place if the bio-cryptosystem has been defeated to do so. Generally, biometric cryptosystems are classified into three systems: key release, key generation and key binding. Core requirements of these systems are revocability, key variety, secure key management and performance. The revocability and key variety characteristics mean that various keys can be reissued for various applications by the same biometrics in case of compromise while secure key management indicates that the key should not be stored at some location. The performance refers to the ability of a biometric cryptosystem to distinguish amongst the genuine users. An incorrect biometric key should be produced when the imposter's biometric template is presented to the system. Overall, the performance of biometric cryptosystems was on average acceptable since the researchers utilised ECC and the pre-processing, the alignment and noise reduction processes, in order to eliminate the biometric variabilities and to raise the accuracy of FAR and FRR. However, caution must be taken when those systems are directed towards implementing transparent multibiometric cryptosystems. The accuracy of FAR and FRR can be degraded as the biometric signals will be very noisy in case of non-intrusive acquisition. In addition to this, the selection of particular biometrics plays the vital role in maximising the performance in practise due to their nature. In particular, the fingerprint and the iris biometrics are very popular as they are quite consistent over time. Moreover, they can be easily converted in to binary representation.

Biometric key release might be appropriate to deal with the cloud storage security. Although this approach does not adequately fulfil the biometric cryptography requirements (see section 2.8) in terms of security, further countermeasures are required to reinforce the security characteristics. Certainly, the key release approach offers certain positive aspects in terms of performance and usability, but there is always the possibility of attacking the stored information in which the secret key can be released. Additionally, the biometric-based key release system does not support the revocability characteristic in case of biometric compromise. Nevertheless, the same secret keys in biometric key release would be liberated at all times without any variances as they are not directly derived from the biometric features. As Uludag et al. (2004) and Kalsoom and Ziauddin (2012) illustrate, the biometric key release system cannot cope with attacks on the sensitive stored information, the biometric template and the cryptographic key. Consequently, when the biometric template is leaked, there could be no point for updating the cryptographic key. This is because the key release approach does not provide cancellable biometric templates. Authors in the key release research area have justified the usage of cryptographic algorithms, such as DES and RSA, to protect the stored information (Soutar et al. 1998); however, these algorithms still use secret keys which must be stored. Recently, Karovaliya et al. (2015) have sought to resolve the issues of cryptographic key management-based biometric identifier through releasing a key via one-time password technique into mobile phones. However, the storage of biometric features with the purpose of recognising the legitimate user is again considered a significant concern vis-a-vis the attacks of imposters. What is more, the one-time password could reflect security and usability issues where the dependence on persons is existing because the code is still in the possession of the user – one-time password

authenticates the presence of that code only not the genuine user. Furthermore, the proposed system still needs to use a cryptographic algorithm with a secret key for secure communication. The context of this research could take advantage the capability of private storage organizations, such as the hard disks to protect the important information within the biometric-based key release. In addition, the cancellable biometric approach can be employed to support the revocation property.

Biometric key generation approaches can be classified into direct and indirect key generation. The former can generate biometric key from the reference template and regenerate it from the test template if necessary. The latter, however, drives helper data/secure sketches from the source template aiding in achieving security applications via the target template. Direct key generation entirely supports the secure key management characteristic. Nonetheless, the reliance upon it for improving the cloud storage security seems to be not suitable solution in terms of usability because of the difficulty of regenerating the same biometric key directly from biometrics throughout. That is, this approach is likely to be inapplicable for encryption and decryption aims as if one bit within the generated key is changed over time, the file will not be decrypted. What is more, direct key generation cannot reissue other keys in case of compromise because it does not store helper data. This opinion is confirmed by Rathgeb and Uhl (2011) with the claim that biometrics perhaps cannot provide robust features to consistently generate long and updatable keys in the absence of public data storage. Further attempts by Hoque et al. (2005), Sheng et al. (2008) and Atah and Howells (2009) improved the extraction of the same biometric key each time directly from biometrics. Their experimental results, however,

were not highly optimistic to replicate these attempts for boosting a more secure and usable bio-cryptographic framework.

Biometric key binding and indirect key generation, in which the key is bound with the template and recovered through successful matching, or created by storing public data from the template, could not introduce a promising solution for protecting the cloud data transparently. According to Uludag et al. (2004), there is no capacity to apply the biometric fuzzy matching specifically in the encryption-decryption scope for particular reasons. One of these reasons is that the biometric features are often noisy and inconsistent over time; therefore, eradicating their variations can be considered very difficult in the encryption-decryption scope. Another reason is that the fuzzy matching of biometrics may not be applied in the encrypted domain as it could be difficult to engineer a meaningful similarity metric in the encrypted representation. This refers that successful fuzzy matching, either it is fuzzy commitment, fuzzy extractor, fuzzy sketch, bio-hashing, or fuzzy vault (i.e. fuzzy approaches) might be inapplicable when the biometric cryptosystem is implemented within a non-intrusive mode for encryption and decryption purposes. This is because the variances in the intra-sample will be too high. As a result, the similarity between the source and query biometric templates cannot be calculated – especially to those biometrics of multiple dimensional features such as the face (Feng and Yuen 2012). Additionally, fuzzy commitment, fuzzy sketch, bio-hashing, and specific fuzzy extractor approaches present a security issue in applying the verification process via hash function. The integrity checking of hash function between the stored hash and the other one at the time of authentication decides whether or not the biometric key is valid for cryptographic applications. Using hash function for verification might

possibly lead to serious security breaches representing in compromising the stored hash, and thus will minimise the security of biometric key management. This view is supported by Stevens (2012) who illustrated that hash functions, such as SHA-0, SHA-1, MD4, and MD5 are not collision-free. This means that it could transform two messages into the same hash values; therefore, security issues will probably take place by masquerade attack. As such, strong cryptographic hash function should be selected carefully to overcome such attacks. For revocation goals, some fuzzy approaches can cancel the compromised instances to reissue other versions as the researchers generated irreversible or locked biometric versions in which the source templates cannot be expected by imposters (Song et al. 2008, Juels and Wattenberg 1999). Nevertheless, other approaches combined the arguably simple passwords with the biometric templates to easily revoke the hacked credentials and reissue new ones. In contrast, this combination will impact the security management in negative aspect because of the need to use the traditional passwords. This opinion is confirmed by Chang (2012) with the claim that there is a negative influence upon the passwords that are used to generate cryptographic keys through password guessing attacks.

In particular, fuzzy vault may not be efficient and effective to be applied within transparent multiple biometric approaches. A possible reason is that the polynomial reconstruction problem alongside the chaff points generation within the traditional fuzzy vault approach can commonly complicate the pre-processing methods for removing the biometric variances (Nandakumar et al. 2007). As a result, the difficulty of applying noise reduction steps might be exacerbated in the case of directing the fuzzy vault towards a non-intrusive bio-cryptosystem development as the biometric

variances will be highly increased. Additionally, the entire complex process of fuzzy vault construction is the creation of the chaff points. These points are used to conceal limited biometric features (Nguyen et al. 2013). Therefore, exploiting limited biometric features in constructing the fuzzy vault may hinder the application of mature, comprehensive and robust transparent multibiometric cryptosystem where the feature vector is expected to be extended. Accordingly, further research would be required to manifest the capability of fuzzy vault approach to achieve non-intrusive bio-cryptography using transparent biometrics.

From the aspect of entropy, this factor is associated with the uniqueness of the biometric modalities, and in particularly evaluates the number of possible feature combinations for the biometric cryptosystem. In simple terms, the entropy factor predicts how difficult a given biometric features would be against brute force cracking. Whenever the entropy of biometric features is increased, the biometric cryptosystem will be strong enough against the efforts which are required to leak the biometric key by adversaries. Consequently, the entropy is a very important factor that should be taken into account when developing and investigating a bio-cryptographic approach. Some of previous research discussed the entropy concept, and evaluated their bio-cryptosystems with regard to entropy. However, other studies did not take the entropy evaluation into consideration (Hoque et al., 2008, Atah and Howells, 2009, Jagadeesan and Duraiswamy, 2010, Chang, 2012, Abuguba et al., 2015). What is more, for the authors who took the entropy into account, they apparently measured the possible values in which a biometric feature vector can have on the presumption that all values had equal probability. Specifically, the majority of contributions tend to reduce the intra-person variances, aimed at coping with the fuzzy matching by

disregarding significant biometric features. As a result, this would probably affect the entropy factor leading to minimise the number of combinations from the feature vector. Accordingly, the biometric cryptosystem will be vulnerable to brute force attack.

Interestingly, the literature review illustrated that the approach of biometric key generation-based recognition can reflect promising and reasonable indication towards developing a novel bio-cryptosystem. This approach exploits a neural network technique to recognise the secret key depending on fresh biometric features and training parameters which are determined and stored at the enrolment stage (Chang, 2012). Such an approach can be possibly adopted to generate a constant and non-intrusive bio-crypto key from transparent biometrics for ensuring data privacy within cloud storage on a timely basis.

3.7 Conclusion

Numerous bio-cryptographic approaches are reviewed and analysed in order to explore the effective solution for coping with the security and usability issues of cloud storage. Essentially, there are three biometric cryptosystems: biometric key release, biometric key generation, and biometric key binding. These systems have sought to cope with a number of serious challenges, such as dealing with the hacked biometric templates, establishing bio-cryptographic keys, and overthrowing the need of using the traditional password. Some researchers succeeded to resolve these issues and to achieve secure solutions. In particular, some papers illustrated that considerable biometric features could be removed in order to eliminate the biometric variances. However, this will probably influence the biometric entropy concept which is

considered very vital factor for a bio-cryptosystem. Consequently, there is a crucial need in order to improve these approaches – specifically making them as secure and usable as possible.

There are a number of approaches that can be exploited for a non-intrusive bio-crypto key generation. At this particular stage of research, it is unclear which approach would be feasible for employing a transparent biometric of more variable features in which a robust entropy and a good accuracy would be accomplished. There is no strong evidence showing that there is a single approach better than another. However, the analysis of the prior research illustrates that Chang's work (Chang, 2012) could reflect a tangible indicator for developing a non-intrusive bio-cryptosystem in a more secure fashion. In Chang's approach, the bio-crypto key is not directly derived from biometrics; thereby, the fuzzy key generation is achieved without disregarding considerable biometric features. As such, it can possibly reinforce the biometric entropy with acceptable accuracy to resist the potential brute force attacks. Although Chang's research reflects a good way for ciphering and deciphering data, the application of single biometric modality, in particular, the keystroke dynamics approach will definitely affect the system with regard to security and accuracy. Therefore, Chang's approach would be taken forward with the aim of enhancing the existing weaknesses and investigating the potential of generating a bio-crypto key from transparent biometrics to develop a viable innovative encryption framework for cloud-based storage.

Chapter Four: Investigation into Transparent Bio-Cryptography

4.1 Introduction

It is evident from Chapter 3 that the prior research of biometric cryptography has presented various approaches to overcome the issues of biometric and encryption. However, there are still specific flaws associated with the security of the resultant bio-crypto key and its usability in practice. Significant research consulted in this project tends to eradicate the biometric variabilities by ignoring considerable biometric features to elevate the system performance. This clearly would impact the biometric entropy factor resulting in reducing the number of possible feature combinations, thus exposing the bio-cryptographic approach within cloud storage to brute force attack. From the usability standpoint, the application of bio-cryptographic keys within cloud-based storage currently poses significant inconveniences throughout. The subscribers will have to present their biometric credentials intrusively each time a file needs to be encrypted or decrypted. This will consequently lead to cumbersome and inconvenient issues while using the cloud storage service each time. The research area of transparent biometric approach offers the opportunity to eliminate the usability issues associated with traditional biometric cryptosystems – potentially enabling more usable and secure cryptography. However, the use of transparent biometrics would likely increase the variability of feature vectors thereby exacerbating the same issue that has always existed for bio-cryptography solutions.

The approach of a biometric key generation based on pattern recognition presented by Chang (2012) could present the potential towards employing transparent biometrics to cope with security and usability issues of cloud storage. Despite this, there are specific issues within Chang's contribution that have to be overcome. That is, Chang's approach is only applied upon the conventional keystroke biometric in which the features, particularly the ones collected intrusively from typing a simple password, will be insufficient for improving cloud-based storage in terms of security and usability. The researcher also did not take into consideration the fact of maximizing the feature vector length with the purpose of strengthening the biometric entropy and combating the potential brute force attack. Furthermore, the correlation between the key size (e.g. 128-bit, 256-bit, 512-bit, etc.) and the accuracy of reproducing the intended key has not been experimentally examined within Chang's research.

It is clear from the above arguments that there is a necessity to resolve the existing weaknesses Chang's work in terms of security and usability in practice. Accordingly, this chapter presents an innovative bio-cryptographic approach based on Chang's scheme to enhance cloud storage service using a number of transparent biometric modalities. In addition, it takes into account the maximum number of possible feature combinations to support the biometric entropy factor in resisting potential brute force attacks. Such an approach needs to be empirically investigated from different perspectives (i.e. security, accuracy and usability) to evaluate its practicality. As a result, three experiments are designed to investigate the effectiveness of the approach. Prior to thoroughly presenting the methodological approach of each

investigation, the core proposed approach is described in the following section in order to appreciate the need for conducting those experiments.

4.2 A Novel Bio-Cryptographic Approach

The proposed approach seeks to develop a convenient and secure user-oriented cryptographic solution to further protect the data privacy of cloud storage by the subscribers themselves. This approach handles the shortcomings of the password login and removes the usability issues of the third-party cryptographic applications. According to Chang (2012), the pattern classification can be exploited to generate a multi binary output as a key using the live biometric features and helper data that is specified and stored on enrolment. As such, the novel approach applies a transparent biometric technique to create repeatable bio-crypto key on the fly via a pattern recognition approach without storing sensitive data (i.e. biometrics and key). In the machine learning field, a long binary key can be established via employing the multi-label classification problem. One of the methods to technically solve the problem of multi-label classification is an adapted algorithm approach. Backpropagation neural network is widely used to solve complex problems in pattern classification and achieved good performance (Chang, 2012). Therefore, the approach of backpropagation algorithm is adapted for generating a long binary key from a transparent biometric technique. The innovative bio-crypto key generation scheme is depicted in Figure 4.1.

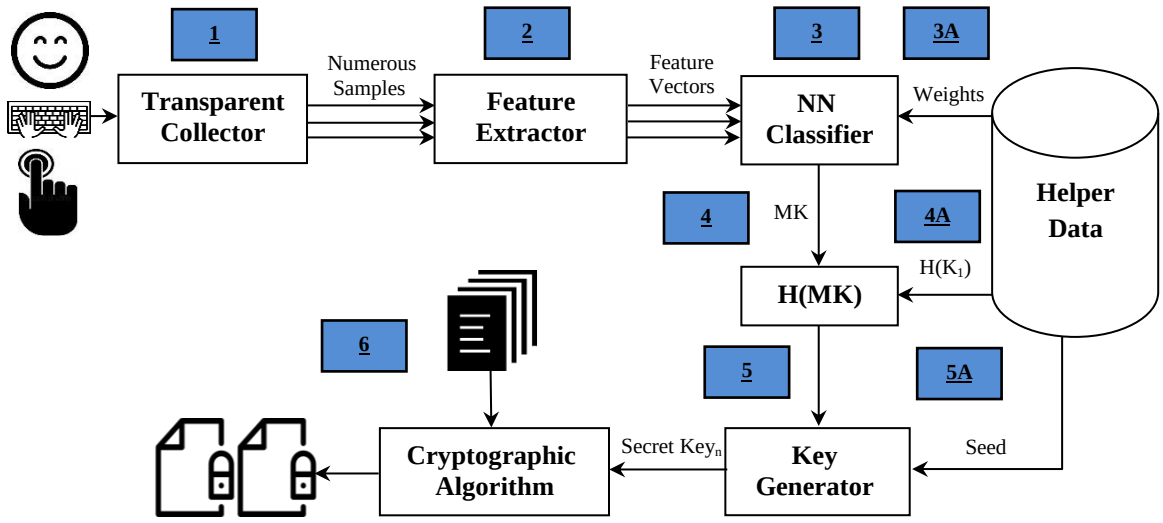


Figure 4. 1 The Innovative Bio-Crypto Key Generation Scheme

In the presented approach, a **Transparent Collector (1)** will at first capture a number of biometric samples such as fingerprint, face, and keystrokes, and subsequently a **Feature Extractor (2)** will take out the biometric features. As such, the reference features would be trained by backpropagation neural network (i.e. **NN Classifier (3)**) to identify a multi-label random key. Then, only the weights in addition to the hashed key on training are stored as a **Helper Data (3A)** to reconfigure the network once again when generating the same key (master key) on encryption/decryption via the fresh features. To this end, a fully-interconnected 3-layer feed forward neural network is configured including input, hidden and output layers. Fundamentally, the input layer is the biometric feature vector, and the output layer is the desired binary key – the master key can be generated on training by using any random key generation approach. Each neuron/node at the hidden and output layer is connected to a bias and equipped to an activation/transfer function according to the goal of the neural network. In backpropagation, the transfer function at the hidden layer either can be tan-sigmoid or log-sigmoid where the former produces [-1, 1] values, and the latter [0, 1] values. In addition, the output layer commonly utilizes any linear

activation function to produce limited outputs of small ranges. As the target of the neural network is a binary random key of a k-bit length, the activation function of log-sigmoid is set at the output layer to produce a binary secret key and the tan-sigmoid at the hidden layer. As a result, the 1's labels would be ranked higher than the 0's labels reliant upon a specific threshold. Of course, the master key will be correct if its hash value (4) equals to the stored hashed key on enrolment (4A). With numerous samples being acquired in a non-intrusive fashion, this verification procedure would allow to trade-off the FRR against FAR to generate the correct key. Within minute window of time, it is spontaneously possible to capture 6 samples. By applying the key verification, only one key is needed to be correct per minute. If the valid key is produced via one successful sample within that time window, the innovative approach would be effective even with 5 samples being rejected. Consequently, 1-minute key generation process is adopted to tackle the high error caused by the transparent collection. The last aspect of the approach is that an individual file seed (5A) stored also as a helper data is entered alongside with the generated master key on encryption into a random Key Generator (5) to produce document keys. Eventually, each document key is used to seamlessly encrypt/decrypt each file within the cloud storage using a sophisticated Cryptographic Algorithm (6) such as AES.

4.3 Research Methodology

As a result of experiencing security and usability problems caused by poor or cumbersome credentials within cryptography, the generation of a constant repeatable bio-crypto key from transparent biometric is investigated. The cryptographic key creation using transparent biometrics can be very challenging as the non-intrusive collection of a sample will result in a higher degree of biometric

variations. In accordance with this, a number of research questions are addressed to be experimentally investigated by the following:

- What is the reliability of regenerating key material from a transparent biometric approach?
- What are the potential factors of classification that might affect the key generation performance?
- What is the correlation between the key size generation and the accuracy of reproducing the required key through biometric features?

Therefore, three experiments were developed to be carried out with the purpose of exploring the derived research questions as below:

Experiment 1 – An investigation into transparent bio-crypto key generation: a baseline set of experiments to investigate how reliable the proposed bio-cryptographic approach at generating a timely constant and non-intrusive key via classification from transparent biometric modalities.

Experiment 2 – An Investigation into improving key generation performance: a series of experiments upon each biometric modality to investigate the factors that could impact classification with a view to enhancing the key generation effectiveness.

Experiment 3 – An investigation into generating different key sizes through features: a set of experiments upon each biometric modality to determine the correlation between the key size (e.g. 128-bit, 256-bit, 512-bit, … etc.) and the accuracy of reproducing the intended key by the biometric features.

The above experiments are all related to each other where the outcome of the first experiment is fed to the second one and the latter to the third experiment. Once experiment one explores the reliability of generating a bio-crypto key from each selective transparent biometric modality, experiment two will seek to enhance the key generation performance by modifying the classification factors of a backpropagation neural network. Experiment three accordingly would investigate the correlation between the key length (e.g. 128-bit, 256-bit, 512-bit, … etc.) and the accuracy of generating the desired key based upon the superior classification approach which is determined by experiment two. Figure 4.2 summaries the methodological approach of the experiments as follows:

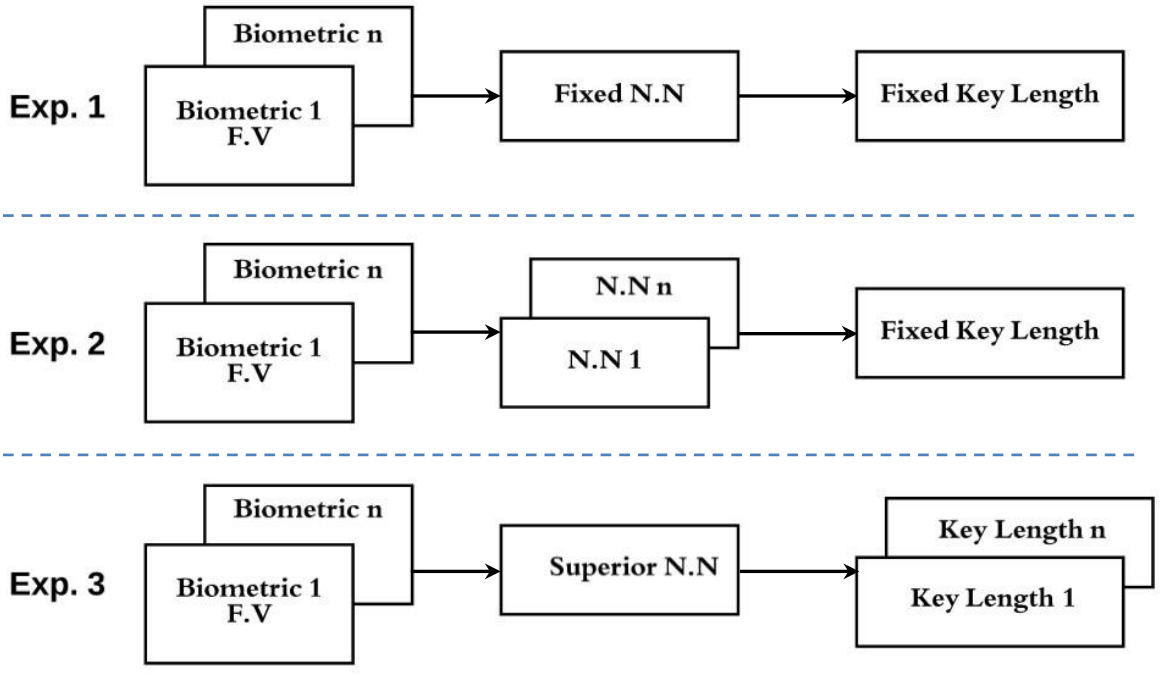


Figure 4. 2 The Methodological Approach of the Experiments

4.3.1 Datasets

With a view to validating the core contribution of this research, functional and practical biometric data collected in a transparent and unconstrained manner was required. Such a collection, however, will produce a more variable feature vector thus complicating the same problem of which bio-cryptography is always struggling to cope with. As a result, a well-established feature extraction technique was needed to tackle those variances. Commercial algorithms can offer an effective approach in extracting a fairly consistent feature vector over time. These algorithms are designed in a reliable way after years of research and development in biometrics. Being a commercial algorithm also includes a number of image analysis and pre-processing methods by which the best quality of feature vector can be obtained (NEUROTECHNOLOGY, 2016).

To the best of the author's knowledge, whilst some commercial vendors enable the developers and integrators to take advantage the entire biometric system only, few of them break the system down into components, such as feature extractor and matcher for a sought goal. In addition to this, these components can be separately utilized in particular within fingerprint and face technologies (physiological biometric modalities only). On the other hand, a behavioural biometric modality needs also to be considered in this research to explore the potential contribution of generating a bio-crypto key from such an approach. Selecting a variety of physiological and behavioural modalities would offer a better insight about the system performance in practice and its potential improvement further on. Accordingly, three biometric modalities including physiological (fingerprint and face), and behavioural (keystrokes

analysis) were selected to determine their contribution within the proposed bio-cryptographic approach.

Collecting biometric data of a sufficient number of users over an appropriate number of samples in a transparent fashion was required. This perspective would reveal a tangible indication with regard to the potential of applying the proposed approach in the real world. The acquisition of transparent biometric samples, however, is not an easy task. Having done an exhaustive search of biometric datasets online, the majority of them was captured in very controlled conditions where they are typical for normal biometric evaluation and do not reflect the characteristics of transparent biometric data. These characteristics represent the acquisition of biometric samples in a non-intrusive manner without the user's explicit interaction in order to incorporate a realistic range of biometric variances. There was a choice to make about whether to carry out a process of capturing transparent biometric data or ultimately employ available datasets that could probably represent the transparent sample collection. Some authors contributed to the research area of biometrics in presenting real biometric datasets consisting of a reasonable reflective population. They also took into consideration the fact of incorporating various biometric variations for experimenting the possibility of developing sophisticated security applications in reality. Using such datasets in these specifications can reflect the sample collection in a transparent and unconstrained manner. As such, it is believed that there was no apparent need to pragmatically undertake the data collection phase in this study.

In seeking the datasets in which a considerable population of biometric samples were captured in a fairly noisy fashion, Yin et al., (2011) introduced a realistic database (SDUMLA-HMT) for a various range of biometric modalities, such as

fingerprint and face. This database was considered very appropriate for the research area of unconstrained biometric in real world environments where diverse conditions of biometric variabilities were incorporated (Yin et al., 2011). As such, the fingerprint and the face datasets of the SDUMLA-HMT were adopted for experimentation. On the other hand, given the absence of the keystroke dataset within the SDUMLA-HMT, another one was needed to suit the others in terms of variability and popularity. Having checked a number of available keystroke datasets online with regards to their protocol for data acquisition, a realistic scenario for typing was set in GREY dataset (Giot et al., 2009) thus triggering real biometric variances to occur. Therefore, this dataset was satisfactorily adopted in order to reflect the non-intrusive sample collection as close as possible. The details of the above-mentioned datasets are described in the following subsections starting with the physiological biometrics of fingerprint and face and then turning into the behavioural biometric of keystrokes analysis.

4.3.1.1 Fingerprint Dataset

The SDUMLA-HMT fingerprint dataset comprised 106 respondents. Each one presented 6 fingers of both hands (i.e. thumb, index and middle fingers) to acquire 8 samples per finger via 5 different scanners – totally 5 sub-datasets (Yin et al., 2011). In a real-time scenario, an application would often capture a single finger for security purposes; therefore, only the index finger (the typical instant) was taken into consideration in this research. The fingerprint dataset which was collected by the AES2501 swipe scanner, was arguably considered more variable than the other datasets since the picture sample size clearly differed in various swiping processes (i.e. there is no fixed size for fingerprint images). Furthermore, while checking some

participants' samples of the same dataset, they were obviously collected by placing a finger in an irregular manner on the capture device, see Figure 4.3. The AES2501 dataset was accordingly employed in this research for experimentation purposes. Figure 4.3 shows some samples of the fingerprint dataset acquired by the AES2501 swipe scanner as below:

Figure 4. 3 Sample Images of Fingerprint Dataset Captured via the AES2501 Scanner (Yin et al., 2011)

Table 4.1 depicts the core fundamental properties of the adopted dataset as follows:

Table 4. 1 The Characteristics of the Fingerprint Dataset

Number of Users	106
Number of Samples	8 Samples for a Finger
Image Size	Not Fixed
Image Type	256 Gray-Level
Image File Format	Bmb

Once the fingerprint dataset was selected and identified in line with the context of this research, a feature extraction approach was needed to extract distinctive values for the biometric key generation process. With the purpose of extracting reliable fingerprint features, a commercial algorithm was sought in order to take advantage of consistent features over time. Having checked a number of commercial technologies, Neurotechnology fingerprint software development technology (NEUROTECHNOLOGY, 2016) was one of the commercial biometric venders that can individually perform the feature extraction process. The feature extraction

approach of Neurotechnology forms up the features in a fashion that allows researchers to understand what a feature dose truly really mean in addition to its start and end in order to experimentally determine if it is useful or not. Therefore, this approach would offer the opportunity of evaluating the biometric entropy based on the actual feature values. The extracted feature vector is also compatible with biometric standards, such as ISO/IEC and ANSI/INCITS standards. Furthermore, Neurotechnology fingerprint demonstrated reliable results in significant evaluations and competitions (e.g. National Institute of Standards & Technology, and Fingerprint Vendor Technology Evaluation for the US Department of Justice) (NEUROTECHNOLOGY, 2016). This confirms that Neurotechnology supplies effective algorithms and credible software development technologies for developing sophisticated secure information technology solutions. Accordingly, the feature extraction algorithm of Neurotechnology fingerprint technique was utilised in this research to extract the fingerprint features. It is worth noting that Neurotechnology does not disclose any details concerning the applied proprietary algorithms in terms of how the fingerprint features are extracted.

Having implemented the feature extraction approach of Neurotechnology, the features of fingerprint minutiae were extracted from 102 users whereas they cannot be extracted from the other four participants due to considerable biometric variabilities. As a result, the feature extraction approach of Neurotechnology flailed to obtain the feature vector for those users. Further, each minutiae-based feature has a set of six values. The feature set of fingerprint minutiae is described in Table 4.2 as below:

Table 4. 2 The Description of Fingerprint Features

Feature Set	Description	Standard Range
Minutiae X	The X coordinate of where this feature is on the fingerprint sample	0-dimenssion of the vertical pixels
Minutiae Y	The Y coordinate of where this feature dose exist upon a sample	0-dimenssion of the horizontal pixels
Minutiae Angle	The angle between the horizontal axis and the direction of the fingerprint minutiae	0-360
Minutiae Quality	A value which determines how bad/good the quality of the fingerprint minutiae (i.e. the higher the value, the better the quality of the minutia. If quality of the minutia is unknown, it must be set to zero)	0-100
Ridge Density	The fingerprint ridge count corresponding to a defined fingerprint area	0-255
Curvature	The level of a ridge near minutia	0-255

The following figure 4.4 explains the minutiae-based feature set – it is worth noting that Neurotechnology does not disclose any illustrations about the features of ridge density and the curvature as they are their own intellectual property:

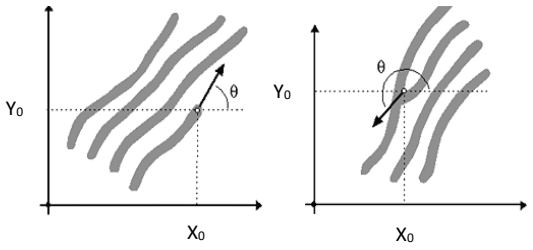


Figure 4. 4 Fingerprint Minutiae-Based Feature (Bansal et al., 2011)

Given samples of different quality, a number of features' sets were varied from one sample to another. As such, presenting various number of minutiae points reflects an indicator about the quality of the feature vector. Even with the commercial approach, there is a difference across acquired samples of participants in terms of

how many minutiae points were extracted in the meantime. As a consequence, this argument demonstrates that the selected fingerprint dataset included high degree of variability amongst samples which could represent a transparent fingerprint collection. Figure 4.5 shows the difference amongst a number of a user's samples as follows:

Figure 4. 5 The Difference Minutiae amongst a Number of a User's Samples (Yin et al., 2011)

In light of the varying minutiae points from one sample to another, it would be improper to compare one point from the sample 1, for instance, with another point from sample 2 of the same user as they are unaligned points. Therefore, there is an apparent need to apply an alignment approach in order to obtain the same number of features among samples across all individuals – thus legitimizing like for like comparison. The problem of fingerprint alignment via the minutiae axises fundamentally falls into a 2D point pattern matching. Accordingly, an approach was implemented with the aim of aligning the minutiae points to be into a common version reliant upon the distance between two point sets of features.

4.3.1.2 Face Dataset

The SDUMLA-HMT face dataset included 106 users, where literally 84 facial samples were gathered from each respondent (Yin et al., 2011). Table 4.3 lists the main attributes of the adopted dataset as follows:

Table 4. 3 The properties of Face Dataset

Number of Users	106
Number of Samples	84 Samples per User
Number of Images	8904 Images
Image Size	640×480 Pixels
Image Type	Colour
Image File Format	Bmb
Total Size of Dataset	8.8 GB

During data collection, a variety of facial variances (i.e. poses, lightings, face expressions and accessories) were captured to reflect a real-life scenario. Therefore, this dataset was considered very challenging in the research area of unconstrained face recognition. As such, it would be evidently appropriate for experimenting the transparent bio-cryptographic approach using the face biometric. With regard to the pose condition, three instances of pose (i.e. looking forward, upward, and downward) were incorporated. Then, 7 samples were taken for each instance to totally obtain 21 samples from each participant as shown in Figure 4.6.

Figure 4. 6 Face Pose Samples (Yin et al., 2011)

For facial expressions, four expressions were identified including smiling, surprising, frowning and closing eyes. Then, 7 samples were captured for each expression thus resulting in 28 samples of face expressions as depicted in Figure 4.7:

Figure 4. 7 Facial Expression Variations (Yin et al., 2011)

From the accessories perspective, a pair of glasses and a hat were utilised as two instances of accessories to collect 14 samples from each individual (7 samples for each instance) as illustrated in Figure 4.8.

Figure 4. 8 Facial Accessories (Yin et al., 2011)

Regarding the lighting variances, three lamps were set to make different lighting angles. Accordingly, 7 samples were captured by illuminating a single lamp only each time - thereby obtaining totally 21 samples of different lighting conditions. The face samples of different illuminations are shown in Figure 4.9 (Yin et al., 2011).

Figure 4. 9 Face Samples of Different Illuminations (Yin et al., 2011)

The facial software commercial technology of Luxand has been considered an advanced technology within the research area of face recognition as it has made significant contributions within academic research and development during the recent years (Luxand, 2016). In particular, Luxand components were applied and referenced in over 200 papers published in renowned scientific journals. Therefore, Luxand was adopted in this research to extract the discriminative facial features where the coordinates of 70 interpretable feature points (including eyes, eyebrows, mouth, nose and face contours) can be detected from a face sample (Luxand, 2016). The features in this manner present a tangible indication regarding how reliable the biometric entropy would be against the potential brute force attack. That is, the biometric entropy can be assessed reliant upon the actual feature values. In addition, each facial feature would be formed up individually in a fashion that determines its start and end in order to experimentally explore if they it is effective or not. Having taken out the coordinates of 70 facial feature points via the feature extraction approach of Luxand, the feature vector extracted for 50-83 samples of 105 users only owing to the huge facial variabilities. As a result, the feature extraction approach of Luxand flailed to obtain the feature vector for those samples. The description of the coordinates of 70 facial feature points is illustrated in Figure 4.10 as below:

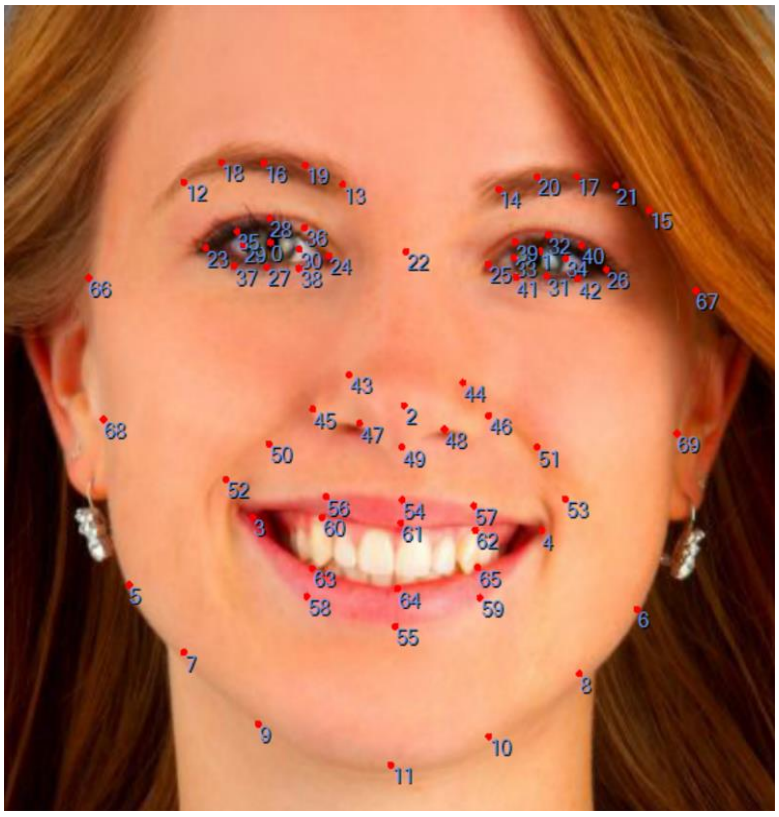


Figure 4. 10 Facial Feature Points (Luxand, 2016)

4.3.1.3 Keystroke Analysis Dataset

On the whole, the GREYC keystroke dataset included 133 users with different number of samples for each (Giot et al., 2009). Table 4.4 illustrates the essential characteristics of the GREYC keystrokes dataset:

Table 4. 4 The Attributes of Keystrokes Dataset

Number of Users	133
Duration of Collection	2 Months
Number of Sessions	5 (1-2 Sessions per Week)
Number of Keyboards	2
Time of Typing Passwords	12 Times per Session
Number of Samples	12-60 samples
Password	greyc laboratory
Number of Total Features	60

Whilst 100 participants presented the highest equal number of samples (i.e. 60 samples per user), the other 33 users had a smaller number of samples – where only 12 samples were captured. Due to the need for sufficient samples, only 100 users having the 60 samples were selected. The features of the keystroke dynamics approach were determined at the same time of data collection; therefore, four keystroke features were identified as follows:

- Difference between two press actions
- Difference between two release actions
- Difference between one press and one release actions
- Difference between one release and one press actions

The keystroke dynamic features were assembled during two months across five sessions, where one or two sessions were performed per week. Prior to gathering features, a data collector asked the participants to log in to the system to practise how to type a uniform password (i.e. greyc laboratory). During the collection stage, the volunteers were asked to type the password 6 times by using two different keyboards in order to assemble realistic features acquired in a real-time scenario with variances. That is, 12 samples were gathered per session, and totally 60 samples per user were obtained during five sessions (Giot et al., 2009).

4.3.2 Investigation into Transparent Bio-Crypto Key Generation

The primary purpose of this experiment was to investigate the reliability of generating a non-intrusive bio-crypto key material on a timely basis by using a number of transparent biometric approaches. Those transparent biometric approaches were physiological modalities (fingerprint, face) and behavioural modality (keystroke dynamics). The challenge represents the ability to generate a reproducible secret

biometric key using a transparent biometric which is inherently noisier than its conventional counterpart. In a non-intrusive biometric mode, the samples would be collected in a different way in comparison with the traditional biometric mode – where there is no obvious interaction between the capture device and the person. This would present more variable biometric feature vector leading to a higher error. As a consequence, there is an essential need to explore how reliable the approach was at generating a transparent bio-crypto key over time to be employed for a seamless cryptographic framework.

In accordance with the proposed bio-cryptographic approach discussed in 5.2, a number of methods were applied to carry out various tasks, such as feature extraction, and the generation of the neural network. These methods are briefly outlined in the following:

- **Feature Extraction:** extracts the feature vectors of fingerprint and face by using commercial algorithms, whereas the features of the keystroke actions are determined at the same time of data collection.
- **Data Manipulation:** normalises the biometric features into the range of 0-1. As Snelick et al., (2003) illustrated that it is necessary to normalise input features into the same range as the output in order to reduce the complexity of the resulting target and the performance.
- **Dataset Splitting:** divides the samples of the selective datasets into two groups: the first was used for training the neural network classifier, and the other was utilised to validate the performance of the classifier.

- **Neural Network Configuration:** configures a fully-interconnected 3-layers feedforward neural network structure by setting its parameters in terms of inputs, hidden neurons, and outputs.
- **Helper Data Construction:** constructs and stores important data (i.e. the weights of the neural network – disadvantageous to attackers) at the time of training in order to generate a repeatable secret key on time.
- **Bio-Crypto Key Generation:** generates bio-crypto keys on training to facilitate in constructing helper data and on validation by using the live biometric features and the stored helper data.
- **Evaluation:** evaluates the performance of the neural network classifier by calculating the accuracy of generating a consistent key each time.

In this experiment, each biometric modality was individually experimented in order to determine the reliability of generating a transparent bio-crypto key from each modality. Having extracted and normalised the biometric features, the classification of a dataset should be fundamentally developed on training data and then applied to test data. Consequently, a split-sample approach was needed to divide a dataset into two groups: registration group and key generation group. The former (registration group) included a number of reference samples for building a user profile and training the classifier to construct an indispensable public/helper data to generate a key in the meantime. However, the latter (key generation group) comprised the fresh samples that were arguably acquired in a non-intrusive and unconstrained fashion to validate the performance of the classifier in generating a constant bio-crypto key on a timely basis. In line with a standard methodology, the splitting approach of 50/50 was performed upon the selective datasets aiming at

dividing each one into registration and key generation groups in order to provide an equal amount of data for each group - thus leading to a realistic evaluation.

With a view to measuring the system accuracy, two metrics of error at rejecting the valid user or accepting a forger to create the required key are considered (i.e. False Rejection Rate (FRR) and False Acceptance Rate (FAR)). Thus, an individual was considered a legitimate user, while all the others were assumed the adversaries who were targeting the valid key of the legitimate user. This process was performed consecutively with the aim of assuring that all individuals had the opportunity to be treated as genuine users, and the results averaged across the population. Table 4.5 demonstrates the genuine user's samples against imposters' samples during the registration and key generation stages reliant on the selected splitting data approach for all modalities.

Table 4. 5 Experimental Settings of the First Investigation

Modality	No. of Users	No. of Samples	Splitting Sample Approach	Registration	Key Generation
Fingerprint	102	8	50/50	Genuine=4 Imposters=404	Genuine=4 Imposters=404
Face	105	50-83	50/50	Genuine=42 Imposters=4368	Genuine=41 Imposters=4264
Keystrokes	100	60	50/50	Genuine=30 Imposters=2970	Genuine=30 Imposters=2970

A backpropagation neural network is widely used to solve complex problems in pattern classification and achieved good performance (Chang, 2012). The Feed Forward Multi-Layer Back-Propagation (FF-MLBP) neural network was consequently employed for key generation/identification. As Hagan et al., (1996) illustrated, $(\text{Input} + \text{Output})^{1/2}$ neurons would be used in the hidden layer of the neural network in order to achieve acceptable classification performance, where the input

represents the number of elementary features for each biometric modality and the output represents the desired key. With regard to the iterations/epochs number of the neural network, a set of tests were performed to determine a satisfactory figure for an effective classification (i.e. the number of epochs was 1000). Table 4.6 accordingly shows the parameters of the FF-MLBP neural network.

Table 4. 6 FF-MLBP Classification Parameters

Modality	Futures/Input	Output (Key)	Hidden	Weights
Fingerprint	516	256	386	200204
Face	140	256	198	28372
Keystrokes	60	256	158	10052

For helper data construction, two target random keys were created and labelled for the feature vectors of the reference samples at the time of training, one for the legitimate user and another (inversed the first) for the remaining presumed imposters with the aim of evaluating the FAR and FRR rates later on validation. A random programming function was exploited to create these keys by seeding the subject number within a dataset. Table 4.7 reveals the creation of a number of genuine users' keys against the imposters' key dependent upon the subject number in a dataset.

Table 4. 7 Generated Keys on Enrolment

Subject No.	Genuine User Key (256-bit)	Imposters' Key (256-bit)
0	1 1 0 0 1 1 1 1 1 0 0 1 0 1 1 0 0 1 0 0 0 1 0 0 1 0 0 0 1 0 0 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 1 1 0 0 1 0 0 1 1 1 0 1 0 0 1 1 0 1 1 1 1 0 0 1 1 0 0 0 0 1 1 1 0 1 0 0 1 1 0 1 1 0 1 0 0 1 1 1 0 1 1 0 1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 1 1 1 1 1 1 0 0 1 1 0 0 1 0 0 0 1 1 1 1 1 0 1 0 1 0 0 0 1 1 1 0 1 0 0 0 0 1 0 1 0 1 1 1 0 0 0 0 0 0 1 0 0 1 0 0 1 1 0 0 0 1 0 0 0 0 1 0 0 0 0 1 1 0 0 1 1 1 0 0 0 0 0 1 1 1 0 1 1 0 0 0 1 1 0 1 1 1 0 1 1 1 1 0 1 0 0 1 0 1 0 1 1 0 1 1 0 1 1 1 0 1 0 1 0 1 1 1 1 1 1 0 0 1 1	0 0 1 1 0 0 0 0 0 1 1 0 1 0 0 1 1 0 1 1 1 0 1 1 0 1 1 1 0 1 1 1 0 0 0 0 0 1 1 1 0 0 0 0 0 0 1 0 0 1 1 0 1 1 0 0 0 1 0 1 1 0 0 1 0 0 0 0 1 1 0 0 1 1 1 1 0 0 0 1 0 1 1 0 0 1 0 0 1 0 1 1 0 0 0 1 0 0 1 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0 1 1 0 0 1 1 0 1 1 1 0 0 0 0 0 1 0 1 0 1 1 1 0 0 0 1 0 1 1 1 1 0 1 0 1 0 0 0 1 1 1 1 1 1 0 1 1 0 1 1 0 0 1 1 1 0 1 1 1 1 0 1 1 1 1 0 0 1 1 0 0 0 1 1 1 1 1 0 0 0 1 0 0 1 1 1 0 0 1 0 0 0 1 0 0 0 0 1 0 1 1 0 1 0 1 0 0 1 0 0 1 0 0 0 1 0 1 0 1 0 0 0 0 0 0 1 1 0 0
1	0 1 1 0 0 1 0 1 0 1 0 1 1 1 1 1 1 1 0 0 0 0 1 1 1 0 0 0 1 1 1 0 0 1 0 0 1 1 1 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 1 0 1 1 1 0 1 1 0 0 0 1 1 1 1 0 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 1 0 0 1 0 0 1 1 0 1 1 1 1 0 0 0 1 0 1 0 1 0 0 0 1 1 0 0 1 1 1 1 1 1 0 0 0 0 1 1 1 1 0 1 0 1 1 1 1 0 1 0 1 1 0 0 1 1 0 0 0 0 1 0 0 1 1 0 0 1 1 1 1 0 1 1 1 0 0 1 1 0 0 0 1 1 1 0 0 0 1 0 1 0 1 1 0 0 1 0 1 1 1 1 0 1 1 0 1 0 0 1 1 1 1 1 1 1 1 0 1 0 1 0 1 1 1 1 0 0 0 0 0 0 1 1 1 1 0 1 1 1 0 0 1 0 1 1 1 0 0 0 0 1 0	1 0 0 1 1 0 1 0 1 0 1 0 0 0 0 0 0 0 1 1 1 1 0 0 0 1 1 1 0 0 0 1 1 0 1 1 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1 1 1 1 1 0 0 0 0 0 1 1 1 0 1 0 0 0 1 0 0 1 1 1 0 0 0 0 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 0 1 1 0 1 1 0 0 1 0 0 0 0 1 1 1 0 1 0 1 0 1 1 1 0 0 1 1 0 0 0 0 0 0 1 1 1 1 0 0 0 0 1 0 1 0 0 0 0 1 0 1 0 0 1 1 0 0 1 1 1 1 0 1 1 0 0 1 1 0 0 0 0 1 0 0 0 1 1 0 0 1 1 1 0 0 0 1 1 1 0 1 0 1 0 0 1 1 0 1 0 0 0 0 1 0 0 1 0 1 1 0 0 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 1 1 1 1 1 1 0 0 0 0 1 0 0 0 1 1 0 1 0 0 0 1 1 1 1 0 1

Accordingly, the feature vectors were learned by the configured backpropagation in order to identify/recognize the target keys. Afterwards, the training parameters in particular the predetermined weights in recognising the valid key were stored only as a helper data at some location. During the phase of key generation, the helper data (weights) re-configured the neural network structure once again to produce the same established key on training by using the feature vectors of the fresh samples.

The non-intrusive sample collection presents more variable feature vector leading to a higher error. The key generation process, therefore, was designed to create a key

through a fixed period of time. In this case, the correct key can be produced reliant upon the trade-off between the FRR and the FAR. As the encryption/decryption process is undertaken in a transparent fashion, enormous samples can be taken without inconveniencing the user. This allows the system to result in a high tolerable FRR, as long as the genuine user can generate the desired key via at least one successful sample within a predefined period of time. As such, a threshold was precisely determined on the basis which obtained the lowest FAR to ensure a valid key generation. Therefore, a correct secret key is generated once a minute as soon as the genuine user would continuously interact with his/her device. On that basis, the proposed system can non-intrusively collect 6 samples per minute, and one sample at least should correctly create the required key. This argument accordingly interprets that the FRR rate of 83% is fairly acceptable within one minute time of window in generating the non-intrusive key of 256-bit length, and simultaneously the FAR would be as minimal as possible. As such, a number of tests were undertaken upon different threshold ranges (i.e. 1 to 0.01 and 1 to 0.001) in order to specify the best accurate threshold for each user. Empirically, the most successful threshold range for the FAR and FRR values across the whole population was the range from 1 to 0.01. For the experimental purposes, the keys of a genuine user on registration and validation are compared with each other in order to calculate the FRR value. However, the key of an imposter on validation is compared with the key of genuine user on registration to measure the FAR value. The FRR value is evaluated when the genuine user cannot utilise the desired key. As a result, if the system does not create the bio-crypto key to the valid user via his/her features, the false rejection number was counted as 1. Accordingly, the FRR rate is obtained by dividing the total false rejection number by the total endeavours of the genuine user to generate his

key, and then the result is multiplied by 100%. Contrarily, the FAR value is evaluated when the presupposed imposters can illegitimately utilise the genuine user's key. Thus, if the system generates the bio-crypto key to the imposters by their features, the false acceptance number was counted as 1. As such, the total false acceptance number is divided by the entire number of forgers, and then the result is multiplied by 100% in order to calculate the FAR rate.

4.3.3 Investigation into Improving the Key Generation Performance

In view of potential weaknesses from the first investigation, influential aspects upon the key generation process were explored in order to escalate the performance. One of those influential aspects can be the imbalance learning of the legitimacy and illegitimacy instances. Training small normal/legitimate samples (minority class) opposite large abnormal/illegitimate samples (majority class) can impact the effectiveness of the classification approach in generating the desired key. In this case, the applied classification algorithm (FF-MLBP) would tend to recognise all samples as a majority class, and mostly lose the capacity to identify the minority class. With the previous methodical approach, whilst the training data of fingerprint included 4 genuine samples against 404 imposters' samples, the reference data of face contained 42 legitimate samples versus 4368 forgers' samples. On the other hand, the populations of a real-time application-based biometric are quite often organised on valid categories against a small percentage of invalid categories (Chawla et al., 2002). Thus, it is believed that the overall performance of each biometric modality degraded because of using inaccurate or biased parameters in classification. Accordingly, it would be broad to re-examine the previous experiments

depending on acceptable insights in order to determine the potential performance in practice.

In seeking the methods of coping with imbalance learning, oversampling technique can improve the classification performance for most imbalanced distributions (He and Garcia, 2008). According to the prior studies, the oversampling approach is considered an effective solution to handle imbalanced spaces in which categories are unequally represented on the order of 100 to 1, 1000 to 1, or 10000 to 1 (Chawla et al., 2002). Given similar categorical representations within the last experiment, the random oversampling technique was adopted to deal with imbalance training. By applying the principle of this method, random samples from the legitimate set were duplicated to be equal to the illegitimate set. In this investigation, the random oversampling technique treated the imbalanced classes at the training time only; otherwise, it would be unrealistic and unfair to be conducted on validation/key generation. Table 4.8 demonstrates the genuine user's samples against imposters' samples via the split-sample approach of 50/50 during the registration and key generation stages for all modalities.

Table 4. 8 Experimental Settings of the Second Investigation

Modality	No. of Users	No. of Samples	Splitting Sample Approach	Registration	Key Generation
Fingerprint	102	8	50/50	Genuine = 404 Imposters = 404	Genuine = 4 Imposters = 404
Face	105	50-83	50/50	Genuine = 4368 Imposters = 4368	Genuine = 41 Imposters = 4264
Keystrokes	100	60	50/50	Genuine = 2970 Imposters = 2970	Genuine = 30 Imposters = 2970

With the potential of further improving the key generation effectiveness being accomplished via the balance training, another set of experiments were conducted

to determine the influence of the classification parameters upon the FF-MLBP. In essence, these experiments were carried out using the same methods within the last experiment, whereas the classifier parameters (e.g. rounds, number of hidden neurones and number of hidden layers) were modified to determine their effect upon the key generation performance. As such, the input and output layers (i.e. the features and the secret key) were constantly set reliant upon the selective biometric modality, where the feature vector will be different from one modality to another. Then, two neural network structures were configured with different settings: the first with one hidden layer and the other with two hidden layers. The rounds/epochs of both network structures were at the figures of 100, 500, 1000, and 2000. With regard to the one hidden layer network, the hidden layer size of 140 and the other of 280 were examined for each round respectively. However, within the two hidden layers network, the number of neurons comprising each hidden layer of 80-80 and another of 120-120 were undertaken for each epoch successively.

4.3.4 Investigation into Generating Different Key Sizes through Features

Given the potential of generating different bio-crypto keys, the correlation between the key length (e.g. 128-bit, 256-bit, 512-bit, … etc.) and the accuracy of reproducing the intended key by the biometric features was investigated. A challenge could be occurred upon the neural network if varying key lengths are needed to be generated. The more the number of neurons (i.e. the desired key) increases, the greater the likely effort upon the network for producing that key. Biometric variations also play a critical role in degrading a key generation. For instance, 40 features could only produce 64-bit key, whilst 80 biometric features may have the capability to generate

64-bit, 128-bit keys or so. Expanding upon this, a number of experiments were developed and carried out to explore the most effective set of features. Then, those effective features would be applied in creating different keys without undermining the biometric entropy. Had the key generation process established using limited set of features, this would probably affect the entropy factor leading to the minimization of feature vector combinations. As a result, the biometric cryptosystem will be vulnerable to the brute force attack.

Fundamentally, this investigation was performed by incorporating additional methods into the methodological approach of the experiment 2. The superior parameters of the neural network in terms of epochs and hidden nodes which were explored in the experiment 2 were also used within this methodology. With the aim of demonstrating the most effective features, an approach for identifying the most important features (feature ranking approach) was required. Having sought a number of feature ranking approaches, random forest algorithm was amongst the most common classification methods that can be employed for effective feature ranking (Louppe, 2014). Therefore, the random forest approach was adopted in this experiment to identify the more discriminative features.

For feature ranking process, the random forest algorithm can determine the important contribution of each feature in successfully identifying a target label during the time of training only and prioritised them accordingly. Therefore, the binary classification problem (1/0) was employed in order to have a target label for each sample. Accordingly, the target labelling procedure was performed by regarding one participant as genuine (1), and the remaining participants as imposters (0). This procedure was successively repeated to ensure that each individual had the

opportunity of acting as the authorised user. As such, the random forest technique ranked the entire biometric features of the reference samples only for each participant on training.

Technically, the random forest derived feature sets called bagging or bootstrap samples by randomly sampling from a feature vector with potential repetitive instances (Louppe, 2014). This was respectively applied to all users' samples to gather multiple bootstrap samples for each user. Subsequently, a forest of base learning algorithms trained the bagging samples of each user (i.e. decision trees) to predict outcomes. As Oshiro et al., (2012) illustrated, tuning the number of trees over 2000 does not additionally improve the random forest accuracy, and might be worse. Therefore, the number of decision trees in this investigation was 2000; the other parameters of the random forest implementation were set as default. Eventually, the priority of each feature was evaluated amongst all trees by summing the number of splits which include that feature proportionally to the number of samples it splits. Thereby, the features were ordered according to their indispensable role in recognising the target – the more the feature importance, the higher the rank awarded.

Having prioritised the features of each biometric modality, a feature selection procedure was performed across a series of validations to examine the capacity of particular significant features in generating the key using the FF-MLBP network. Thus, the feature selection procedure was set out by selecting the first 20 top features, and thereafter the feature selection was progressively escalated by 20 at each run (i.e. 20, 40, 60, 80, 100, 120, 140 ... etc.). This procedure was individually implemented upon the feature vectors of 516, 140, and 60 for fingerprint, face and

keystrokes respectively. While experimenting a set of features, the key of 256-bit length was set at the output layer. At the same time, the effective entropy of particular feature sets from each biometric modality was evaluated to determine which effective feature set can be applied with a reliable entropy figure for generating different keys. For entropy evaluation of a feature set, a possible value was calculated for each feature from each modality. The possible value is the difference between the maximum and the minimum values a biometric feature can have. Then, the $\log_2$ is taken for the product of those possible values to calculate the entropy of a biometric modality in an effective bitlength.

Having demonstrated the biometric features effectiveness in producing the key of 256-bit, another set of experiments was performed to investigate the accuracy of generating different bio-crypto keys (e.g. 128-bit, 256-bit, 512-bit, … etc.) from each biometric approach individually. Accordingly, the effective biometric features were set at the input layer of the FF-MLBP classifier. Balancing the training classes alongside the superior classification parameters were also adopted. Then, the output layer size was changed at each test to accommodate a different desired key of n-bit length (i.e. 128-bit, 256-bit, and 512-bit). This would demonstrate the capability of the novel approach in producing stronger secret keys. The longer the cryptographic key size is established, the high effort upon the attacker to crack that key.

4.4 Results and Analysis

Having implemented the previous methodological approach, the performance of the transparent biometric key generation via the fingerprint, face and keystrokes biometric modalities is evaluated. The implementation of experiments was

accomplished via python programming language. A number of python programming scripts was written and generated on a Windows 7 Enterprise 64-bit Operating System with Intel Core i5-4310 CPU, 2.7 GHz and 16 GB RAM. The following sections are devoted to presenting and analysing the results of each experiment:

4.4.1 Experiment 1: Transparent Bio-Crypto Key Generation

The experimental results and analysis of this investigation demonstrate how reliable the proposed approach at generating a bio-crypto key material from the contributory biometric techniques in this research. On the whole, the results of this investigation reveal that a fairly reliable repeatable key of 256-bit length can be generated by fingerprint and keystroke dynamics modalities to encrypt/decrypt data in reality. However, the bio-crypto key cannot be reliably produced from the face biometric modality.

It can be observed that the proposed bio-crypto key generation approach presented in section 4.2 can be effective by particularly using the transparent fingerprint modality. Figure 4.11 illustrates the performance of the transparent key creation from fingerprint technique as below:

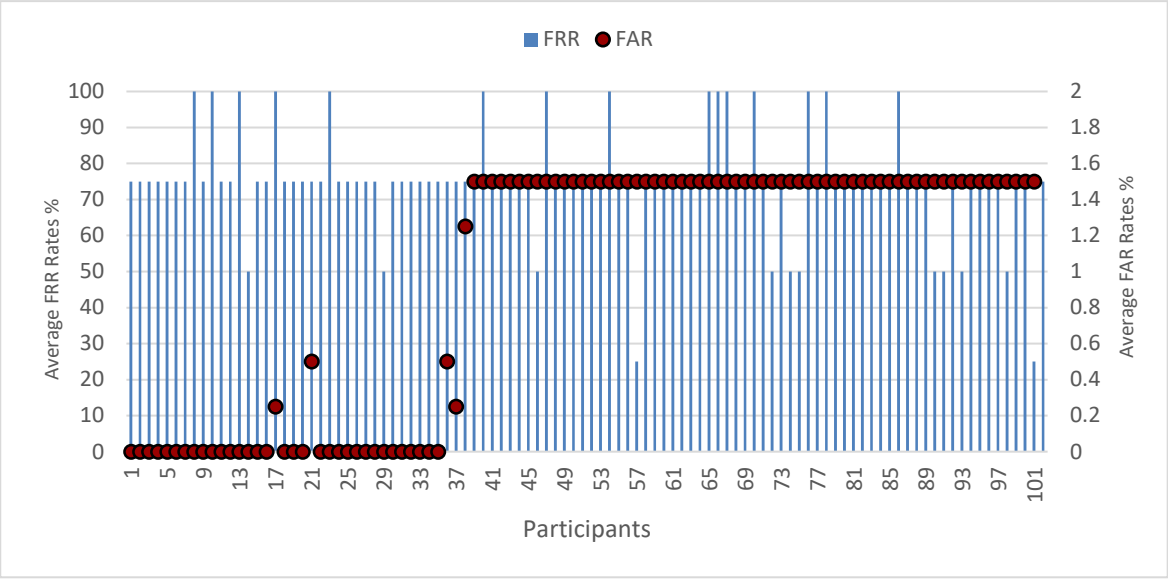


Figure 4. 11 The Performance of the Transparent Key Creation from Fingerprint

It is clear from the chart that the majority of participants generated the key with very limited forgery attempts, where the FRR and FAR rates ranged between 25%-75% and 0%-1.5% respectively. As the 1-minute key generation approach was proposed to establish a key by at least one sample of the collected six ones, the FRR rate of 83% is pretty acceptable in generating the non-intrusive key of 256-bit length. This interprets why the obtained FRR rates (i.e. 25%-75%) were considered fairly acceptable.

On the other hand, well under quarter of the population (15 users) failed to generate the key with a 100% FRR (entirely unacceptable). A possible interpretation about this might be because of training noisy fingerprint samples. As explained in section 4.3.1.1, the fingerprint features were evidently highly variable – where a different number of minutia points were extracted amongst the samples of each user. Furthermore, there was a limitation within the fingerprint dataset in the sense that the number of samples were small - totally 8 samples for each user. By applying the sample splitting approach of 50/50, the entire training set were only 4 reference

samples. It is worth noting from another aspect that while evaluating the performance, the selection of the definitive thresholds between the FAR and FRR rates occupy a discrete range of values rather than a continuous basis. This is due to the limited number of the fresh samples. This could explain why the FAR and FRR results occur in a distinct/separate representation.

With regard to the face biometric, the accuracy of generating the non-intrusive bio-crypto key from the face modality is unexpectedly very poor. Figure 4.12 shows the effectiveness of the transparent key generation approach via the face biometric by the following:

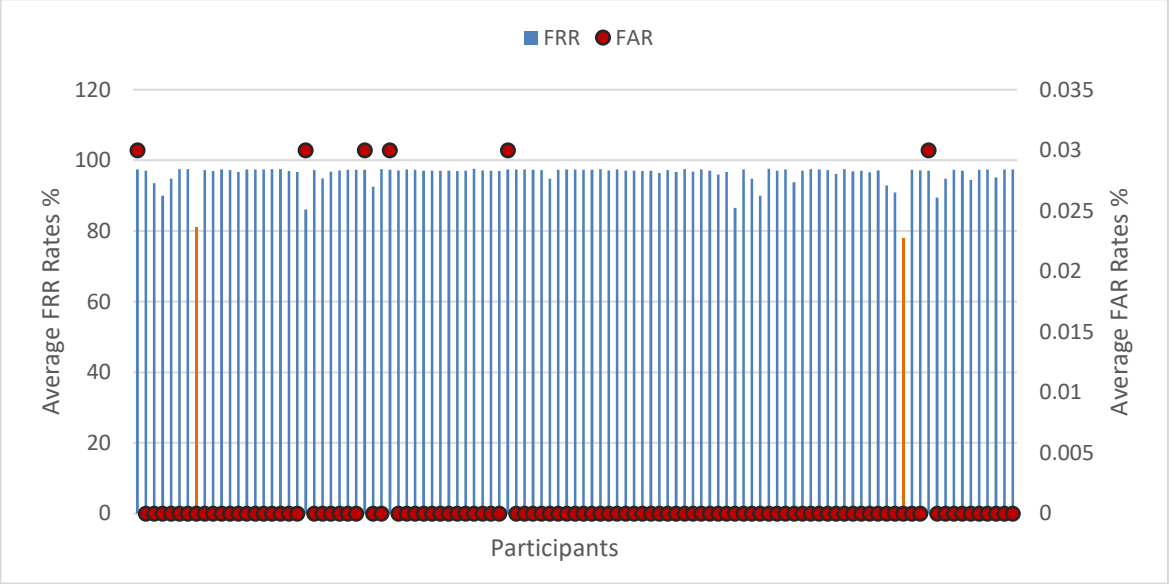


Figure 4. 12 The Effectiveness of the Transparent Key Generation Approach via Face

According to the chart, the whole population (except 2 users) was unable to generate the key of 256-bit length (i.e. FRR was 96.05% on average) – obviously exceeding the acceptable FRR rate of 83%. On the contrary, the significant majority of respondents achieved a minimal figure of FAR (i.e. 0.001), although it is worthless with the accompany FRR rate being accomplished by the same users. This could be

because of training highly inconsistent facial samples – where those samples covered a significant range of face variabilities to reflect various real-life scenarios. In addition to this, the imbalance training classes (42 legitimate samples versus 4368 forgers' samples) could probably have a negative impact upon the key generation performance. This interprets that the classifier tended to generate an invalid key to the minority samples of the legitimate class. Another explanation is that a biased classification might be happened because of using inaccurate number of epochs and/or number of hidden neurons.

The key generation approach by keystroke analysis modality positively accomplishes encouraging accuracy in generating the key of 256-bit length. Figure 4.13 describes the performance of the transparent key generation:

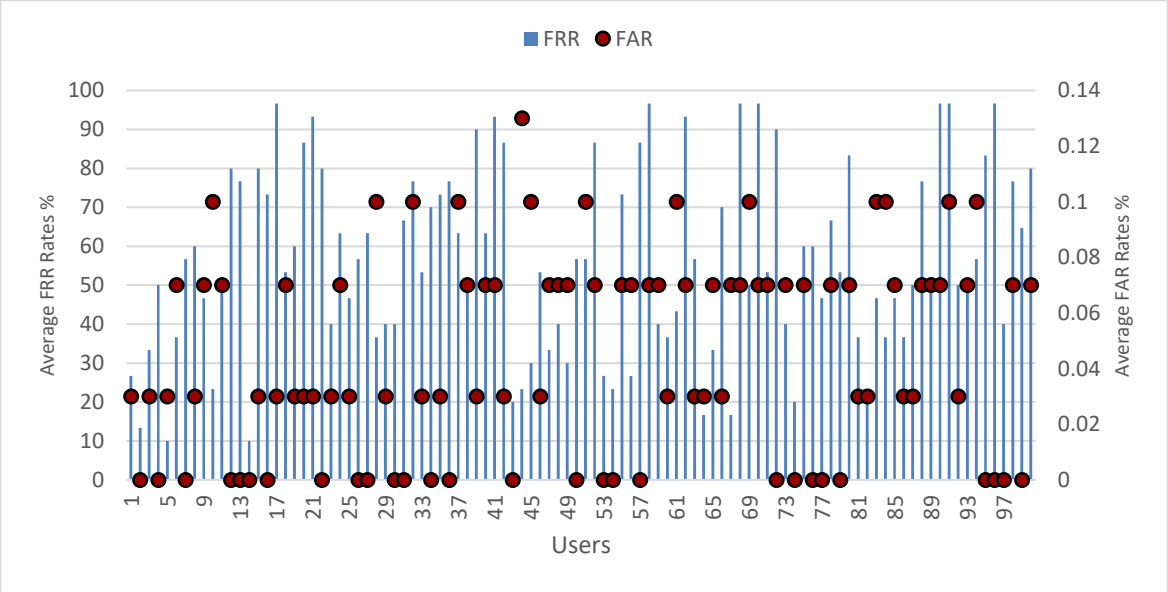


Figure 4. 13 The Performance of the Transparent Key Generation via Keystrokes

It can be seen from the chart that the majority of participants generated the bio-crypto key, where the FAR and FRR figures ranged between 0%-0.13% and 10%-80% respectively. A possible interpretation is that the training of fairly acceptable 30

legitimate samples against 2970 illegitimate samples might be sufficient to accomplish a successful key generation. On the other hand, 91.66% FRR rate (very negative) on average was achieved by well under quarter of the population (18 users) due to the potential of collecting pretty variant keystroke actions within overlapping intervals.

4.4.2 Experiment 2: Improving the Key Generation Performance

The empirical results and analysis of this investigation determine the impact of the imbalance training and the classification parameters upon the key generation accuracy. The experiments' findings generally highlight that the imbalance learning and the improper classification parameters have a negative influence upon the key generation performance. The effectiveness of generating the bio-crypto key (256-bit length) from the selective biometric modalities is clearly improved, when balancing the legitimate and illegitimate spaces on training.

All in all, the performance of the fingerprint key generation is relatively ameliorated as outlined in Figure 4.14.

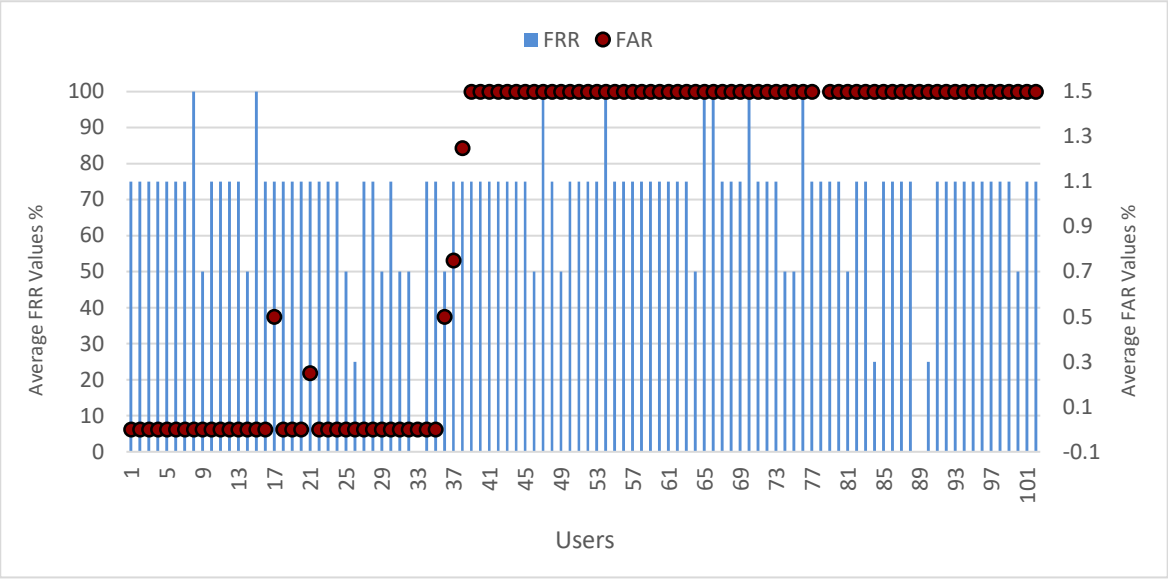


Figure 4. 14 The Performance of the Fingerprint Key Generation

It is obvious from the results that the vast majority of participants achieved acceptable FRR and FAR figures via the balanced training – where the FAR and FRR rates were around 0.9% and 70% respectively. However, only 8 users cannot generate the key with 100% FRR due to the possibility of learning noisy fingerprint samples in addition to the deficiency of the training samples – merely four source samples. In comparison with the imbalance training outcomes, the FAR and FRR results evidently confirm that the balance learning does enhance the performance of the classifier in correctly generating the key of 256-bit length.

The accuracy of generating the key from the face biometric modality is positively improved on balance training as shown in Figure 4.15.

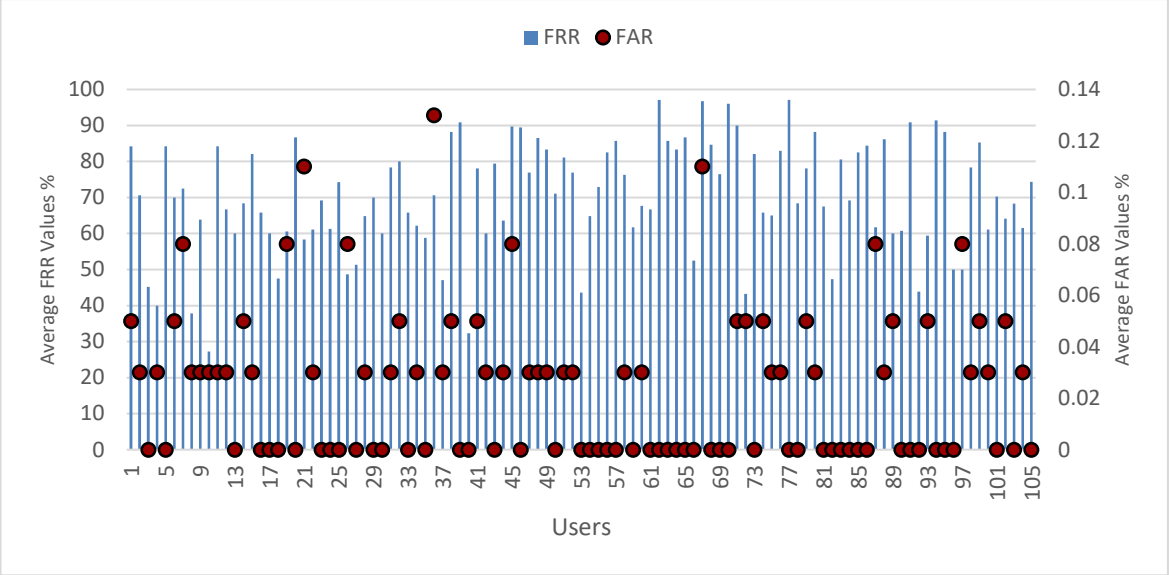


Figure 4. 15 The Accuracy of Generating a Key from the Face Biometric Modality

According to the results, the majority of users succeeded to create the key with 0.02% FAR and 64.6% FRR on average. A possible explanation is that the balanced learning of 4368 valid face samples against 4368 invalid samples can improve the key generation effectiveness. On the contrary, a quarter of the participants did not succeed to generate the key, where unsatisfactory FRR rates were accomplished – higher than 83%. This can be possible owing to the extreme variabilities of the face samples as explained in section 4.3.1.2.

The performance of producing the bio-crypto key of 256-bit length by the keystroke dynamics technique is enhanced on balance training as demonstrated in Figure 4.16.

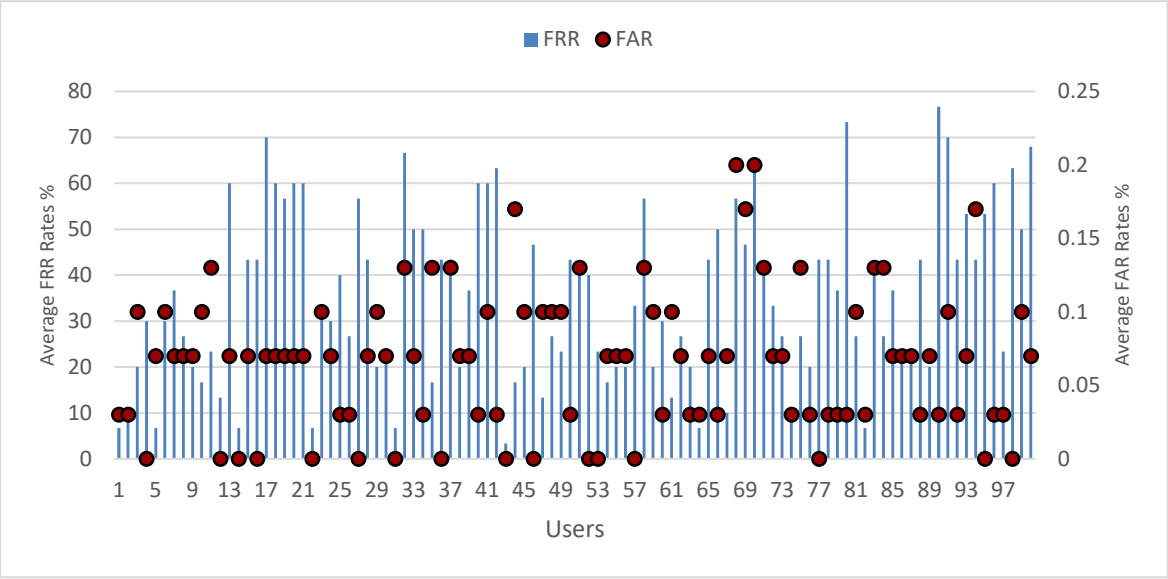


Figure 4. 16 The Performance of Producing the Bio-Crypto Key of 256-Bit Length by the Keystroke

It is evident from the chart that the entire population had the capability to generate the key through the keystrokes biometric modality with a minimal illegitimacy access. The experimental results reported very positive FRR and FAR rates on average which were 35.3% and 0.06% respectively. A possible interpretation is that the accurate classification improved the key generation process overall – especially to those 18 users who failed to create the correct key during the time of training on imbalanced instances. Of course, whenever the FAR and FRR rates decrease, the reliability of the key generation process would be elevated.

The other experiments' findings also show that the key generation performance is overall enhanced by varying the classifier parameters (i.e. number of epochs, number of hidden neurones and hidden layers). The key generation performance of a 256-bit through the fingerprint biometric approach (516 features) using the single and the double hidden layers is depicted in Tables 4.9 and 4.10 as follows:

Table 4. 9 Fingerprint Key Generation Using Single Hidden Layer

Epochs	Hidden Nodes	FAR	FRR
100	140	0.906%	73.04%
500	140	0.953%	67.89%
1000	140	0.975%	65.44%
2000	140	0.958%	62.99%
100	280	0.955%	73.04%
500	280	0.955%	67.16%
1000	280	0.965%	65.2%
2000	280	0.953%	62.25%

Table 4. 10 Fingerprint Key Generation Using Double Hidden Layer

Epochs	Hidden Nodes	FAR	FRR
100	80-80	0.91%	70.1%
500	80-80	0.94%	62.75%
1000	80-80	0.96%	62.01%
2000	80-80	0.94%	59.56%
100	120-120	0.96%	73.53%
500	120-120	0.95%	67.16%
1000	120-120	0.96%	66.67%
2000	120-120	0.94%	64.71%

In Table 4.9, the results reveal that there is apparently no difference in the key generation accuracy using the one hidden layer of 140 and 280 nodes. A possible reason is that the learning based upon a few fingerprint samples (totally 4 samples for each user) could probably need a reasonable number of nodes at the hidden layer. However, as Table 4.10 shown, the two-hidden layer achieved a superior accuracy in creating the biometric key depending upon the fingerprint data. In particular, the double hidden layer of 80-80 outperforms the other layer of 120-120 in generating the secret key of 256-bit. The results of creating a more reliable key using the two hidden layers of 80-80 nodes with 2000 epochs are depicted in Figure 4.17 reliant upon individual FAR and FRR for each user:

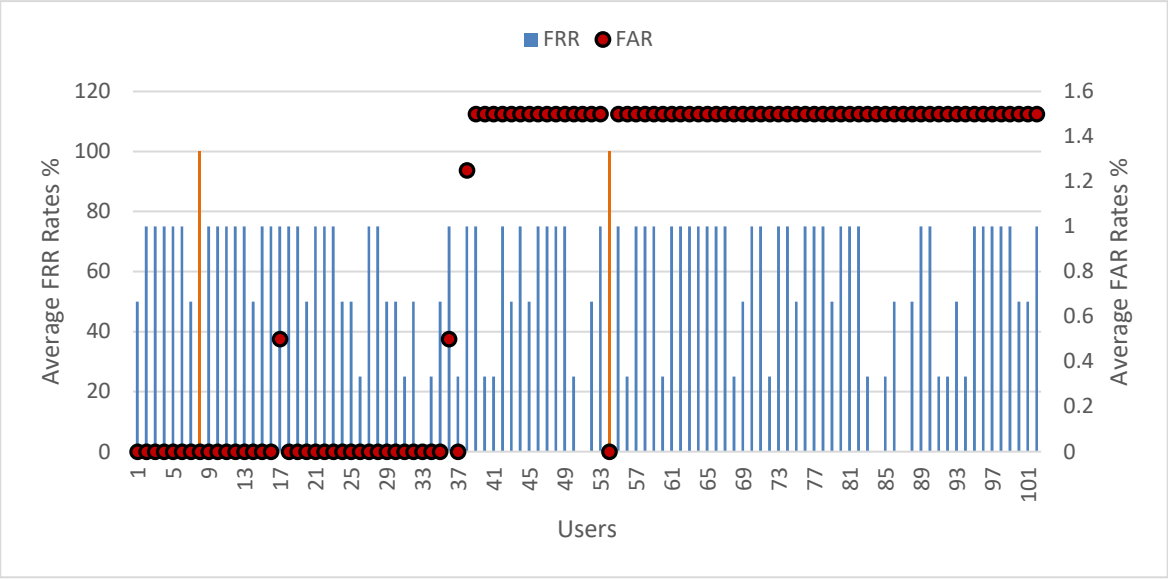


Figure 4. 17 Fingerprint Key Generation Using the Two Hidden Layers of 80-80 Nodes

It is clear from the chart that the whole population (except two participants) was able to create the bio-crypto key of 256-bit, where the FRR value was 59.8% on average - lower than the tolerable FRR figure of 83%. At the same time, the results of this experiment reveal that limited illegitimacy attempts were taken place (i.e. the FAR rate was 0.94%). This might be because the 80-80 neurons within the two hidden layers fit the limited samples of the fingerprint modality to tune the neural network for creating the bio-crypto key. Given the deficiency in the fingerprint samples, this experiment showed that the neural network size does not need to be huge. Furthermore, equalizing the genuine samples versus the adversary samples can have a positive impact in enhancing the key generation performance.

The key creation effectiveness of a 256-bit by the face biometric technique (140 features) using the one and the two hidden layers is presented in Tables 4.11 and 4.12 as below:

Table 4. 11 Face Key Generation Using Single Hidden Layer

Epochs	Hidden Nodes	FAR	FRR
100	140	0.07%	84.4%
500	140	0.08%	76.26%
1000	140	0.09%	69.97%
2000	140	0.14%	67.24%
100	280	0.07%	81.2%
500	280	0.07%	73.92%
1000	280	0.08%	65.81%
2000	280	0.09%	64.89%

Table 4. 12 Face Key Generation Using Double Hidden Layer

Epochs	Hidden Nodes	FAR	FRR
100	80-80	0.02%	86.03%
500	80-80	0.04%	71.09%
1000	80-80	0.06%	65.13%
2000	80-80	0.07%	63.27%
100	120-120	0.02%	83.9%
500	120-120	0.04%	68.74%
1000	120-120	0.06%	62.69%
2000	120-120	0.07%	60.2%

The experimental results reveal that there is an improvement at producing the bio-crypto key using the one hidden layer of 280 nodes over 140 nodes as outlined in Table 4.11. With numerous facial samples being used in those experiments, the 280 nodes within the single hidden layer apparently have the capacity to achieve a better performance. On the other hand, the double hidden layer in this investigation outperforms the single layer in generating the secret key reliant upon the applied facial data. As illustrated in Table 4.12, the double hidden layer of 120-120 accomplished the superiority in producing the bio-crypto key of 256-bit in comparison with the other layer of 80-80. The empirical results of generating a more effective key via the two hidden layers of 120-120 nodes with 1000 epochs are illustrated in Figure 4.18 depending on individual FAR and FRR for each user. Although there were other

experiments relatively achieved more accurate FAR and FRR results than the above-mentioned basis using the two-hidden layer, the number of users that succeeded in generating the key within those experiments is less.

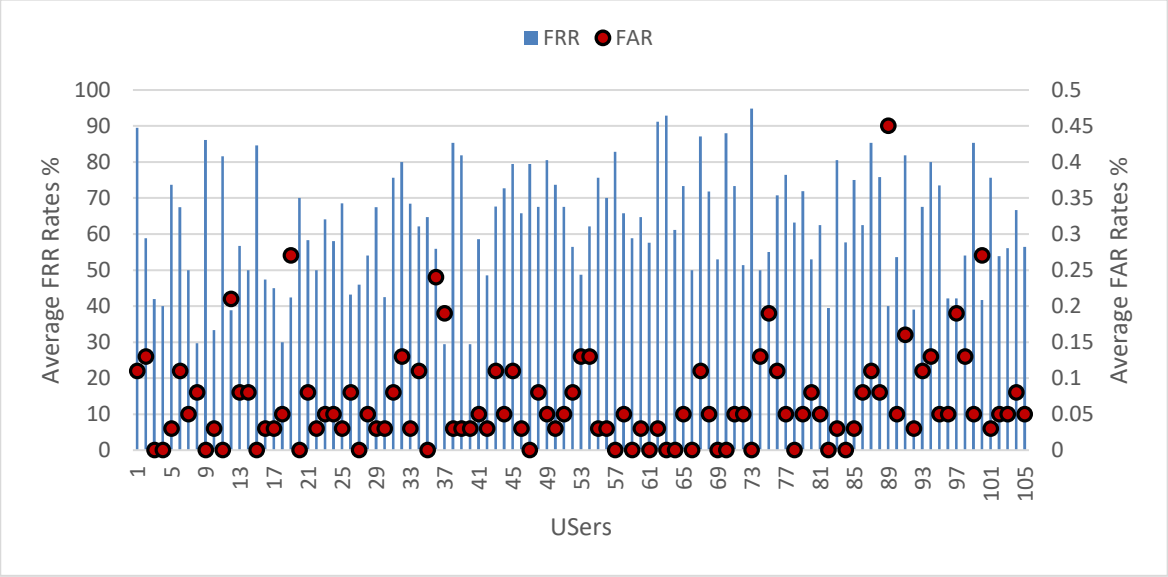


Figure 4. 18 Face Key Generation Using the Two Hidden Layers of 120-120 nodes

It is obvious from the chart that the vast majority of users correctly generated the key of 256-bit length, with 0.06% FAR and 60.2% FRR values on average being positively accomplished. On the other hand, only 9 users failed to generate the bio-crypto key, where FAR and FRR rates were 0.03% and 88.33% (i.e. FRR was greater than 83%) respectively. This could be possible because fairly proper classification parameters can fit the facial data in this experiment to identify/generate the correct key besides the positivity of the balance training.

The key generation accuracy of 256-bit via the keystroke biometric modality (60 features) using the single and the double hidden layers is compared in Tables 4.13 and 4.14 as below:

Table 4. 13 Keystroke Dynamics Key Generation Using Single Hidden Layer

Epochs	Hidden Nodes	FAR	FRR
100	140	0.16%	58.39%
500	140	0.19%	32.48%
1000	140	0.2%	30.45%
2000	140	0.25%	28.61%
100	280	0.29%	53.92%
500	280	0.49%	35.16%
1000	280	0.55%	29.73%
2000	280	0.59%	27.86%

Table 4. 14 Keystroke Dynamics Key Generation Using Double Hidden Layer

Epochs	Hidden Nodes	FAR	FRR
100	80-80	0.05%	38.58%
500	80-80	0.06%	33.27%
1000	80-80	0.06%	33.4%
2000	80-80	0.06%	32.87%
100	120-120	0.06%	36.35%
500	120-120	0.06%	32.4%
1000	120-120	0.06%	32.53%
2000	120-120	0.06%	31.13%

According to the results, using the single hidden layer of 140 nodes based upon the keystrokes data reported pretty good results on the whole. Considering both the FAR and FRR rates from those experiments seem to show a promising reliability and usability. In contrast, the experimental results show that the two-hidden layer outperformed the single layer in generating the cryptographic key. In particular, the experiment of using the double hidden layer of 120-120 nodes with 2000-epoch demonstrated a superior performance, with 0.06% FAR and 31.1% FRR on average. Generalizing the results from this experiment seem to be more effective because a minimal FAR rate is achieved and simultaneously the FRR rate does not exceed the tolerable figure (lower than 83% FRR). The empirical results reveal that all users were capable to create the secret key with a very limited forgery access via the

keystroke analysis technique as depicted in Figure 4.19. A possible interpretation concerning such encouraging results is that expanding the neural network size into 120-120 fits the training keystrokes data – thus decreasing the error rate in terms of the FRR values.

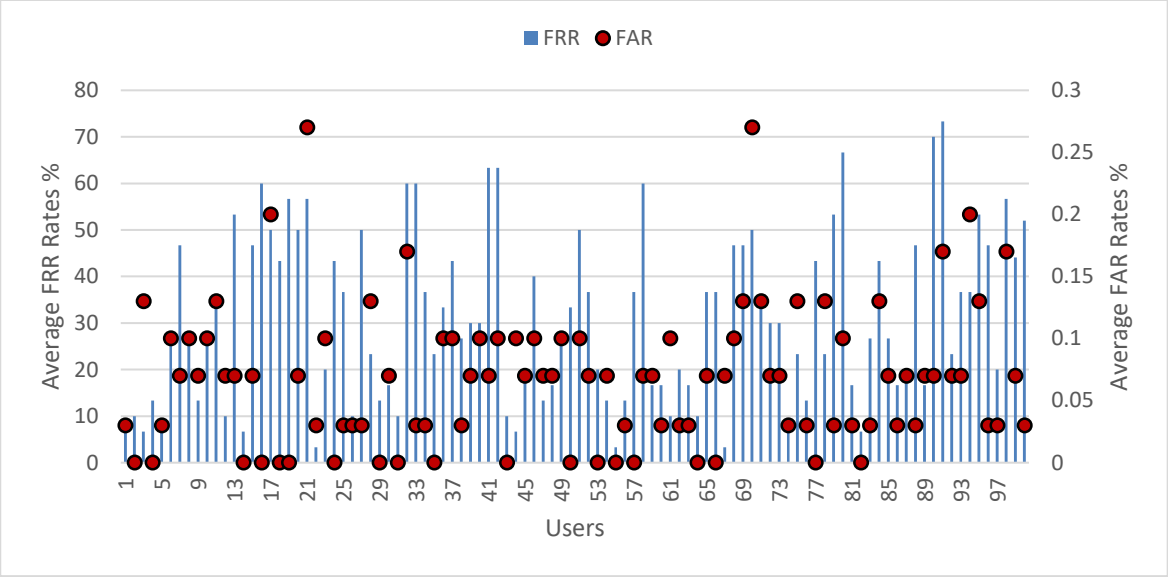


Figure 4. 19 Key Generation Using Double Hidden Layer of 120-120 Nodes via Keystrokes

4.4.3 Experiment 3: Key Length versus Feature Length

The results and analysis of this investigation reveal the correlation between the key length (e.g. 128-bit, 256-bit, 512-bit … etc.) and the accuracy of regenerating the desired key by the biometric features. Prior to undertaking this hypothesis, a set of experiments preliminarily explored the potential effectiveness of the biometric features in generating the bio-crypto key of 256-bit in order to demonstrate a proper trade-off between accuracy and entropy aspects.

Figure 4.20 depicts the key generation performance via examining different ranges of the top-ranked fingerprint features using the random forest technique:

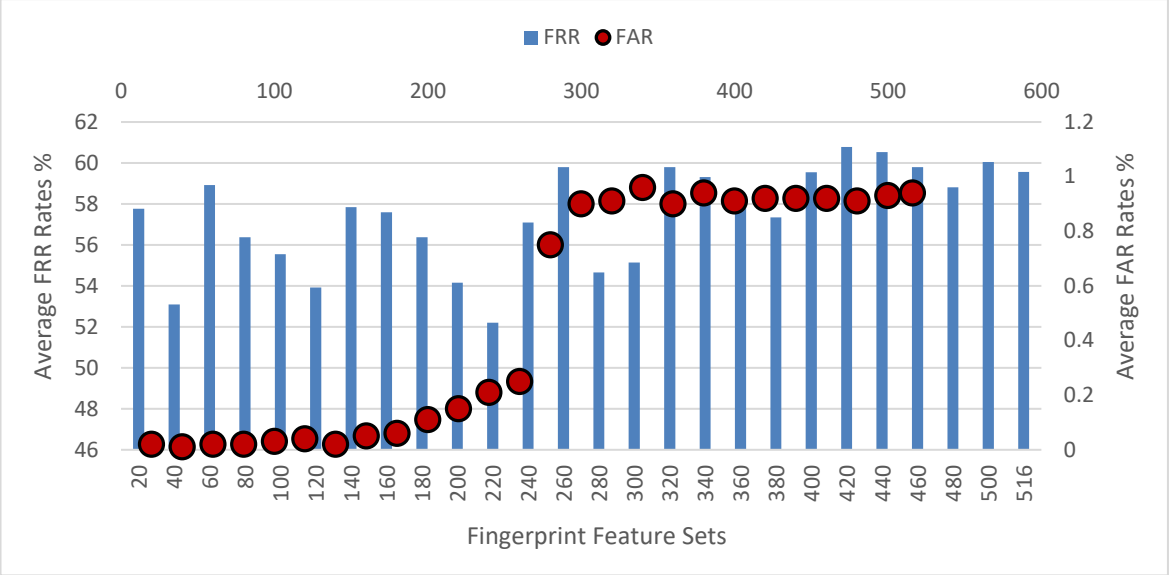


Figure 4. 20 Key Generation via Different Ranges of Top-Ranked Fingerprint Features

According to the chart, the accuracy of the key generation overall does appear to be affected when increasing the top-ranked feature set; especially with the feature sets from 260-516. On the other hand, the fingerprint feature sets of 40, 120, 200 and 220 achieved fairly positive results in terms of accuracy with 53.3% FRR and 0.07% FAR on average. Accordingly, there is an evident improvement in the key generation performance of a 256-bit length. The FFR figure is within the tolerance for this research (lower than 83% FRR), and the FAR rates were overall low.

From a different aspect, the effective entropy of a feature set should be also taken into consideration to reinforce the security of the bio-cryptosystem. Table 4.15 reveals the effective fingerprint entropy for particular feature sets as follows (see section 4.3.4 for entropy evaluation):

Table 4. 15 The Effective Entropy (Bitlength) versus Different Fingerprint Feature Sets

Feature Set	20	120	200	220	512
Effective Entropy	120	290	1498	1652	3817

It is clear from Table 4.15 that adopting only 40 or even 120 feature set is apparently inappropriate for real-life application as they cannot adequately boost the biometric entropy - where the effective bit-length of entropy is 290 and 881 respectively. In contrast, with entropy figures of 1498 and 1652 respectively, the fingerprint feature sets of 200 or 220 might be sufficiently robust against feature guessing attempts. However, it is believed that applying those subsets of features within the bio-cryptosystem would impact the biometric entropy. The 200/220 feature sets would minimize/diminish the number of combination values of the feature vector, thus essentially undermining the entropy factor. On the other hand, the entire feature set of 512 with an effective entropy of 3817 obviously would reinforce the biometric entropy, and can be more reliable in resisting potential brute force attacks. Therefore, in terms of the trade-off between security and accuracy, taking into consideration the performance results via the 516-feature set seems to be more crucial in reinforcing the biometric entropy.

Table 4.16 compares the key generation accuracy through partitioning the top-ranked facial features into a number of different ranges utilizing the random forest approach:

Table 4. 16 Key Generation via Different Ranges of Top-Ranked Face Features

Feature Set	20	40	60	80	100	120	140
FRR	63.15	62.21	62.31	62.26	62.26	62.31	62.69
FAR	0.048	0.054	0.078	0.064	0.08	0.063	0.065

Generally, the experimental results show that the FAR and FRR rates do not reveal a huge difference by increasing the top-ranked feature set. The feature ranking via the random forest technique clearly showed that the small feature vector carries the most discriminative information. With significant facial variabilities being applied in

this research, the extracted features via holistic-based coordinates could not be highly distinctive. The results show that using the most discriminative features from the classification perspective is efficient; however, the biometric entropy should also be considered to combat the brute force attack by incorporating numerous features. Table 4.17 shows the effective bitlength of entropy for particular feature sets as below:

Table 4. 17 The Effective Entropy ((Bitlength) versus Different Face Feature Sets

Feature Set	20	40	60	100	140
Effective Entropy	163	327	491	819	1144

According to the results, the effective entropy of the feature set 40 is 327, whereas the entropy of the feature set 140 is 1144. As such, applying the entire feature vector (140 features) is more desirable to make sure that an attacker would be unable to guess what the notable features are.

Table 4.18 compares the key generation accuracy through partitioning the top-ranked keystroke features into a number of different ranges using the random forest method:

Table 4. 18 Key Generation via Different Ranges of Top-Ranked Keystrokes Features

Feature Set	20	40	60
FRR	33.87	30.17	31.13
FAR	0.078	0.77	0.06

On the whole, the experimental results reveal that the effectiveness of generating the bio-crypto key is very slightly ameliorated when increasing the top-ranked feature subset size. The feature subset of 60 in particular achieved better performance, where limited illegitimacy attempts are demonstrated by the FAR value, and at the

same time the FRR value is less than acceptable figure of 83%. Simultaneously, using the whole feature vector (60 features) with no doubt would resist the potential brute force attack, where the effective bitlength of the feature set 60 is 1504. On the other hand, the results of this experiment overall do not show a significant difference in the key generation accuracy by the keystroke features. A possible explanation might be because of gathering keystroke actions within the same concurrent times.

The previous experimental results confirm that using the entire feature vector across the contributory biometric modalities is crucial, where it supports the security, and would not significantly impact the accuracy. As such, the empirical results of different key lengths creation are presented and analysed based upon adopting the wholly biometric feature vector.

Table 4.19 demonstrates the accuracy of different key lengths generation via the selective biometric approaches (fingerprint, face and keystroke) as follows:

Table 4. 19 The Capacity of Generating Different Key Lengths

Modality	Epochs	Hidden Nodes	Features	Key Size	FAR	FRR
Fingerprint	2000	80-80	516	128-bit	0.938%	58.08%
Fingerprint	2000	80-80	516	256-bit	0.948%	59.56%
Fingerprint	2000	80-80	516	512-bit	0.943%	59.803%
Face	1000	120-120	140	128-bit	0.061%	62.1%
Face	1000	120-120	140	256-bit	0.065%	62.69%
Face	1000	120-120	140	512-bit	0.07%	62.74%
Keystroke	2000	120-120	60	128-bit	0.064%	30.95%
Keystroke	2000	120-120	60	256-bit	0.067%	31.13%
Keystroke	2000	120-120	60	512-bit	0.07%	31.64%

It is apparent from the tabulated results that there is no difference in the key generation performance when creating varying key lengths. This demonstrates that the back-propagation neural network had the capability to reliably map/represent the

biometric data in generating a bio-crypto key. A possible interpretation is that the neural size (i.e. hidden layer size) across all the selective biometric modalities was big enough (see Table 4.19). Thus, the neural network could probably have sufficient memory capacity to represent a particular feature vector quite well for identifying a target bio-crypto key no matter how long that key is. Of course, the longer the bio-crypto key size is generated, the high effort upon the adversary to attack that key.

4.5 Discussion

Generally, the study findings confirm that a robust repeatable cryptographic key of 256-bit length can be positively generated from the selective transparent biometric modalities to encrypt/decrypt the data in reality. The classification conception is exploited to generate constant cryptographic keys through transparent biometric by using the stored classification parameters (weights) and the test features. Consequently, high variations caused by the non-intrusive collection would not overly impact the performance as the key is not directly driven from the noisy features. At the same time, this approach accomplished the necessity of including many features to reinforce the biometric entropy.

With numerous samples being collected transparently in the meantime, the usability aspect would not affect the performance leading to eventually generate the required key as well as would allow the possibility of trading-off the FRR against the FAR. As such, the average FAR was 0.9%, 0.06%, and 0.06% for fingerprint, face, and keystrokes respectively. At the same time, the average FRR value for each selective transparent biometric approach in this study was under the tolerable figure of 83%. On the whole, the key generation performance appears to be impacted by the

imbalance training. In contrast, the effectiveness of the bio-crypto key production is clearly ameliorated on the balanced learning. On imbalance classes, 15 users of the entire population (i.e. 102 individuals) were unable to generate the cryptographic key from the fingerprint modality. On the other hand, the accuracy is evidently improved when balancing the legitimate and illegitimate fingerprint samples, where only 8 users did not succeed to create the key. In terms of generating the transparent bio-crypto key through the face biometric approach reliant upon the imbalance training, the performance is unexpectedly fallen down – where 103 users of the whole population (105 individuals) were incapable to generate the key of a 256-bit (very negative). However, the effectiveness of the key generation based on balancing the facial instances is positively enhanced, where merely 25 users failed to generate the biometric key. Respecting the key generation performance by the keystrokes approach through the imbalanced training, 18 users of the entire population (i.e. 100 individuals) were unable to generate the key. On the contrary, the accuracy is clearly improved when balancing the valid and invalid keystrokes samples, where all users had the capacity to create the bio-crypto key. With further improvement being accomplished via amending the classification parameters, it seems that expanding the neural network size tends to represent the selected biometric data quite well to generate a more effective bio-crypto key. As such, all users (except two) succeeded to generate the key from the fingerprint and keystrokes, whilst only 9 users failed to create the secret key by the face biometric approach. On the other hand, this could not be the case in reality as the biometric data would be different.

The key generation from the keystrokes modality is somewhat more effective than the other modalities (fingerprint and face). It is believed though that the effectiveness

of the keystrokes could be improper to generalize. This opinion is corroborated by Monrose and Rubin, (2000) with the claim that the keystrokes features are inferior compared to the transformational features of fingerprint and face. To clarify, the keystroke features are gathered simply by calculating the time difference between the press actions which are pretty variable. In contrast, the feature of fingerprint and face could be possibly extracted as a geometrical/structural feature or holistic-based coordinates. The former is extracted by engineering certain relations between the attributes besides applying additional transformations to eliminate the variances and derive the most discriminative feature – resulting in reducing the feature vector. The latter, however, is taken out by determining the coordinates of a distinctive location – leading to many features (Monrose and Rubin, 2000, Mohammadzade et al., 2018). As illustrated in the literature review, reducing the features would impact the biometric entropy factor and minimize the number of combination values of the feature vector. In this context, the bio-cryptosystem can be vulnerable to brute force attack. As such, this research adopted the holistic feature extraction within the finger-print and face approaches which can be slightly less effective than the keystrokes to develop a robust trade-off between the security and accuracy.

By incorporating considerable features within the system, there would be a problematic issue to guess them by forgers. The experimental results demonstrate that using the whole feature vector would escalate the number of combination values resulting in boosting the entropy to combat the brute force attacks. The effective entropy for each biometric modality is demonstrated in Table 4.20.

Table 4. 20 The Effective Entropy (Bit-length) from each Biometric

Modality	Fingerprint	Face	Keystrokes
Effective Entropy	3817	1144	1504

The empirical results also interestingly reveal that different secret key lengths can be generated via the back-propagation classifier. This can be due to the possibility of having adequate capacity to map the biometric data to shape a pattern capable of producing a desired key. Creating a longer bio-crypto key would be undoubtedly more resistant vis-à-vis the attacker attempts to crake the cryptographic key.

4.6 Conclusion

This chapter has proposed and demonstrated through experimentation a novel transparent bio-crypto key generation approach to handle the shortcomings of the password login and removes the usability issues of the oriented cryptographic means. The results have empirically shown that a reliable non-intrusive cryptographic key can be generated on the fly without storing it anywhere from the selective transparent modalities. Although the training of imbalanced legitimate and illegitimate classes has experimentally affected the key generation process, the results of balance learning have positively improved the performance in creating a reliable secret key. The experimental results show that the size of the neural network in addition to the number of epochs can influence the key generation performance. With superior key generation performance being achieved by amending the classification parameters, the double hidden layer neural network in this study outperforms the other the single layer configurations. The two-hidden layer can fit/map the experimental biometric data quite well to generalize/model a pattern aiding ultimately in generating the bio-crypto key. There is also an impact upon the

key generation performance when modifying the number of epochs – mostly whenever the number of iterations is increased, the FRR is decreased with slight escalation in the FAR value. As a result, the classification parameters must be chosen carefully to fit the applied biometric data in reality to generate the desired key.

Employing numerous biometric features is desirable in order to support the biometric entropy and to combat guessing biometric feature attempts. What is more, generating the required key as more reliable as possible is also very necessary. Accordingly, the following chapter will seek to investigate a potential effective solution aiming at coping with the above-mentioned implications and eventually reinforcing the security and accuracy aspects.

Chapter Five: Investigation into Transparent Multibiometric Cryptography

5.1 Introduction

Given the desirability to further reinforce the effectiveness of the proposed key generation approach in terms of security, accuracy, and usability, a reliable solution has to be considered. This represents the guarantee of generating the correct key and simultaneously maximizing the biometric entropy against the guessing feature attack. There is no doubt that the application of multiple biometric approaches can aid in generating the correct bio-crypto key reliant upon more than one biometric modality. The problem of unsuccessful key generation could be alleviated by incorporating two or more biometric modalities in order to improve the accuracy and concurrently reinforce the security. For example, those users who cannot create the cryptographic key through the face biometric they might be able to produce it by the fingerprint and/or keystroke dynamics techniques and vice versa. At the same time, this incorporation would reinforce the security aspects in terms of escalating the entropy of biometric features and resisting the forgery attacks. The biometric entropy will be strengthened via accumulating numerous features from all biometric modalities. The spoofing attacks will be also difficult if not impossible where a forger will need to spoof all three biometric samples. From the usability perspective, the multi-biometric approach would additionally reduce the user inconveniences caused by generating the incorrect key. As enormous samples will be collected in a non-intrusive fashion from multiple biometric modalities, the opportunity of creating the correct valid key will be escalated. On the other hand, a multibiometric fusion approach should be effectively applied in order to introduce a constructive manner

rather than destructive. That is, the multibiometric fusion approach must outdo the single biometric technique with regards to security and accuracy. Thereby, the transplant multi-bio-cryptosystem would overcome the illegitimacy and inaccuracy issues owing to the presence of multiple independent pieces of evidence (i.e. applying more than one biometric).

As discussed in 2.5, a source should be provided to fuse some data and thereafter develop a multiple biometric system. Since numerous biometric samples can be non-intrusively collected from more than one biometric modality, the following sources are adopted in this research – where one or set of them is implemented as appropriate:

1. **Multi-Modality Source:** applying a single sample of more than one modality to tackle the weaknesses of some biometric techniques or acquisition devices.
2. **Multi-Sample Source:** implementing multiple inputs of the same modality to have a well-informed identity and to offset the existing samples of low quality.
3. **Hybrid Source:** non-statically applying single or multiple samples from different modalities. This could probably fine-tune the approach in generating the required bio-crypto key, crafting a more multi-layered method.

According to the discussion in 2.5, these samples have to be incorporated effectively at a certain phase (i.e. sensor, feature, matching score, and/or decision level) within the biometric system. For instance, the feature vectors from multiple biometric modalities can be appropriately fused to correctly generate the constant key in a more secure fashion. Another example is that the classification results from multiple modalities could be incorporated with the purpose of successfully enhancing the key generation process. As such, this chapter considers the principle of multibiometric to

combine the fingerprint, face and keystroke dynamics modalities aiming at improving the overall performance.

5.2 Methodological Approach

Having demonstrated the generation of a timely reliable bio-crypto key from the contributing single modalities, the multibiometric fusion was experimented to determine whether the key generation performance would be improved. As referred earlier, a fusion method can take place at any stage within the biometric system. However, there are a number of issues associated with the development of the multibiometric system including the source of information, selective biometric, information fusion, cost-benefit, sequential processes and level of reliability (Nandakumar, 2008). Therefore, there is apparently no conclusive evidence revealing that there is a fusion method overwhelmingly better than another at a specific point (Monwar, 2013). As such, fusion methods were carried out to determine what impact the multibiometric approach would have upon the key generation performance at the feature phase versus the matching phase.

In order to develop a multi-biometric approach, it is necessary to incorporate the samples of the biometric modalities together. Unfortunately, there was a limitation within the size of the fingerprint dataset, where it included the minimum number of samples for each user. As a consequence, a standard procedure was performed by taking into account the lowest common number of samples across all modalities. On the other hand, there would be a significant number of face and keystrokes samples being ruled out aside. Had the methodological approach performed by using the lowest common number of samples only, there would be an unreliable insight about

the multibiometric performance. In order to cope with this issue, another procedure was conducted by using the oversampling technique to duplicate the samples of the fingerprint modality into a reasonable figure. In accordance with this, a number of research questions are derived to be addressed and explored experimentally as follows:

- What is the performance of the multi-biometric key generation approach via feature-level fusion reliant upon:
  - A- Minimum number of samples across all selective modalities (fingerprint, face, and keystrokes)?
  - B- Oversampling technique?
- What is the accuracy of the multibiometric cryptography approach by matching-level fusion dependent on:
  - A- Minimum number of samples amongst all modalities?
  - B- Oversampling method?

Consequently, two fundamental experiments were developed to be performed aiming at resolving the derived research questions as below:

Experiment 1 – An investigation into transparent multi-biometric key generation at the feature phase: set of experiments to explore the potential of improving the key generation accuracy by combining the feature vectors of the applied biometric modalities depending upon appropriate procedures.

Experiment 2 – An investigation into transparent multi-biometric key generation at the matching phase: a number of experiments to investigate the likelihood of elevating the bio-crypto key generation performance by integrating the matching scores from

each classifier being utilized within the individual biometric approach based on an effective fusion technique.

The following subsections describe the nature of the applied multi-biometric dataset with the context of this research, and then turn into illustrating the methodological approach of the experiments.

5.2.1 Datasets

In order to validate the potential of improving the key generation process via the fusion principle, a realistic multi-biometric dataset including the selective transparent modalities was needed. To the best author's knowledge, there is apparently no existing multi-biometric dataset incorporating an adequate range of biometric variations in which the fingerprint, face and keystrokes samples are gathered from the same person. This is a common issue within the topic of this research. With a view to handing this problem, a possible experimentation can be applied via integrating biometric modalities from different datasets; thus, they can belong to the same individual – relying upon so-called virtual user configuration. However, the idea of virtual users is only justifiable in approaches which can be manifested to be independent. From an experimental standpoint, the selected biometric datasets in the previous chapter can be possibly fused to create virtual users as they are arguably independent. Whilst the fingerprint and face datasets were collected from the same participants, the keystrokes dataset was captured from different volunteers. Accordingly, virtual users were configured by using these datasets. With a virtual user being the combination of a user from the fingerprint, face and keystrokes datasets, each individual from the fingerprint dataset was associated with

the others from the face and keystrokes datasets. Combining the subjects in this fashion produces a multi-biometric dataset where the number of the virtual users is equal to the smallest number of users within the individual biometric datasets. As such, the ultimate multi-biometric dataset of 100 virtual users (i.e. each one having his fingerprint, face and keystrokes samples) was configured to be utilized in the experiments.

5.2.2 Investigation into Transparent Multibiometric Key Generation by Feature-Level Fusion

Given the availability of the core biometric identity at the feature extraction stage, the feature-level fusion is sought to investigate the potential for outperformance in generating the secret key of a 256-bit length. The primary objective was to combine the feature vectors amongst the contributing biometric modalities. An investigative issue, however, was encountered during the experimentation. The number of samples in between the fingerprint, face, and keystrokes datasets were unequal (i.e. 8 samples for fingerprint, 50-83 samples for face, and 60 for keystrokes). With the presence of a different number of samples across modalities, two procedures were performed in order to unify them. The first procedure was implemented by setting the minimum number of samples across the selected biometric datasets. As the fingerprint dataset included the minimum number of samples (i.e. 8 samples for each user), the foremost procedure unified the number of samples by randomly picking 8 samples from the other modalities. Therefore, an experiment was conducted to investigate the performance of the key generation by feature level fusion among the fingerprint, face and keystrokes on the basis of setting 8 samples for each user.

On the other hand, another procedure was applied based on oversampling technique via randomly resampling with replacement (i.e. a sample can be duplicated multiple times within a user samples set). Accordingly, the oversampling technique was performed upon the fingerprint dataset only to produce 50 samples; thus, the latter procedure unified the number of samples by randomly picking 50 samples from the other modalities. Regarding the second procedure in which the fingerprint samples were duplicated into 50 sample, it is believed that duplicating the entire fingerprint samples once could lead to an inaccurate validation. So as to cope with this issue, the fingerprint samples were equally divided into training and testing groups and then the sample duplication was performed at each group individually. As such, another experiment was carried out to determine the effectiveness of multibiometric cryptography approach via feature level fusion on the basis of setting 50 samples for each user. The former standard procedure adopted few samples; however, the latter provided the opportunity for eliminating the limitation of applying few samples via including significant biometric samples. Whilst neither one of the procedures would be alone sufficient in resolving the research question, both of them were performed to provide reliable insight about the multibiometric key generation performance. Therefore, it is believed that conducting these approaches only would be appropriate enough to accomplish this investigation. They would rule out the downside of having to duplicate the facial samples as well – the face samples of the entire population ranged between 50-83 samples and the majority of users had only 50 samples. From the feature dimensionality viewpoint, since it is demonstrated that there was no high impact upon the key generation performance by applying the feature selection approach, the entire feature vectors were utilized in this investigation. Accordingly, the feature vectors of the biometric modalities were concatenated in the form of

[$Fingerprint_{FV}$, $Face_{FV}$, $Keystrokes_{FV}$] to obtain the final multi-biometric feature vector.

Similarly to previous experiments of Chapter 4, a number of methods were applied including dataset splitting, neural network configuration, helper data construction, key generation and error evaluation to carry out the feature fusion experiments. The sample-splitting approach of 50/50 was also used in this investigation. The first 50% of samples utilized for developing a user profile and training the classifier to construct a helper data. The other 50% of samples used to test the performance of the classifier in generating a constant bio-crypto key. The balance training demonstrated the capacity of improving the key generation process within single biometric techniques. As such, the experiments would be also carried out reliant upon balancing the valid and invalid instances. The neural network of double hidden layer was also applied in this experiment as particular experiments of Chapter 4 showed that it tends to be more superior in generating the bio-crypto key than the single hidden layer. The experimental settings and the classification parameters of this investigation are presented in Table 5.1 and Table 5.2 as below:

Table 5. 1 Empirical Settings of Feature-Level Fusion

Modalities	No. of Samples	Sample Splitting	Registration	Key Generation
Fingerprint + Face + Keystrokes	8	50/50	Genuine=396 Imposters=396	Genuine=4 Imposters=396
Fingerprint + Face + Keystrokes	50	50/50	Genuine=2475 Imposters=2475	Genuine=25 Imposters=2475

Table 5. 2 Classification Parameters of Feature-Level Fusion

Modality	Futures	Key	Hidden	Weights	Rounds
Fingerprint + Face + Keystrokes	716	256	250-250	349204	1000

The generation of a non-intrusive, continuous and secure cryptographic key is dependent upon the availability of the collected biometric modalities. However, one or more biometric modalities may or may not be present as samples can only be captured transparently if they are available to capture. As such, having performed the biometric fusion of three modalities, other experiments were also implemented based on different combinations of two biometric techniques including fingerprint and face, fingerprint and keystrokes, and face and keystrokes for further analysis.

5.2.3 Investigation into Transparent Multibiometric Key Generation by Matching-Level Fusion

With the capacity for developing a highly modular approach being offered at the biometric matching stage to combine multiple modalities, an investigation was conducted to explore the key generation performance via the matching-level fusion. This experiment was literally carried out by replicating the same methods of the last investigation without using the feature fusion strategy. A matching fusion method aimed to accumulate the scores from each classifier being utilized within the individual biometric modality triggering to triple the outputs. As a result, a normalization technique was needed to rescale the scores into [0, 1]. This technique can be simple sum, median, min, max, or majority voting. However, the security aspect needs more consideration – particular biometric modality should not have much value than another; especially the behavioural biometric (keystroke dynamics). A consideration should be given to the matching fusion method to generate the correct bio-crypto key in a more robust way. As such, the majority voting technique was performed in order to rigorously take the proportional different key generation performance of the contributing biometric modalities into account. In matching

fusion-based majority voting mechanism, the final correct key of a 256-bit will be established if it has been predicted most frequently via the classifiers being used within the single biometric technique (i.e. hard voting). If the prediction of the samples is a correct key from the fingerprint biometric, correct key from the face, and a wrong key from the keystrokes technique, then the final key would be correct. In this case, the matching fusion would have been constructive; otherwise, it will be destructive. Table 5.3 demonstrates the final key generation based on combining the fingerprint, face, and keystrokes techniques at the matching stage using the majority voting mechanism as follows:

Table 5. 3 The Final Key Generation by Majority Voting Mechanism

Generated Key by Fingerprint	Generated Key by Face	Generated Key by Keystrokes	Final Key
Wrong	Wrong	Wrong	Wrong
Wrong	Wrong	Correct	Wrong
Wrong	Correct	Wrong	Wrong
Wrong	Correct	Correct	Correct
Correct	Wrong	Wrong	Wrong
Correct	Wrong	Correct	Correct
Correct	Correct	Wrong	Correct
Correct	Correct	Correct	Correct

The majority voting technique averaged the probabilities of classifiers (i.e. soft voting) for fusing two biometric techniques at the matching stage. One significant challenge that can encounter the matching fusion approach is the manipulation of the score from each individual biometric modality classifier (neural network). However, as each of the three biometric modalities was designed utilizing the same classifier, the scores of the three approaches were already in the same form – enabling their direct use within the majority voting technique. Likewise in the last investigation, a number of methods were applied including sample splitting, neural

network configuration, helper data construction, key generation and error evaluation to carry out the matching fusion experiments. The same empirical settings of the feature level fusion were also used in this investigation. Figure 5.1 describes the matching fusion process of the selective biometric modalities using the majority voting technique as below:

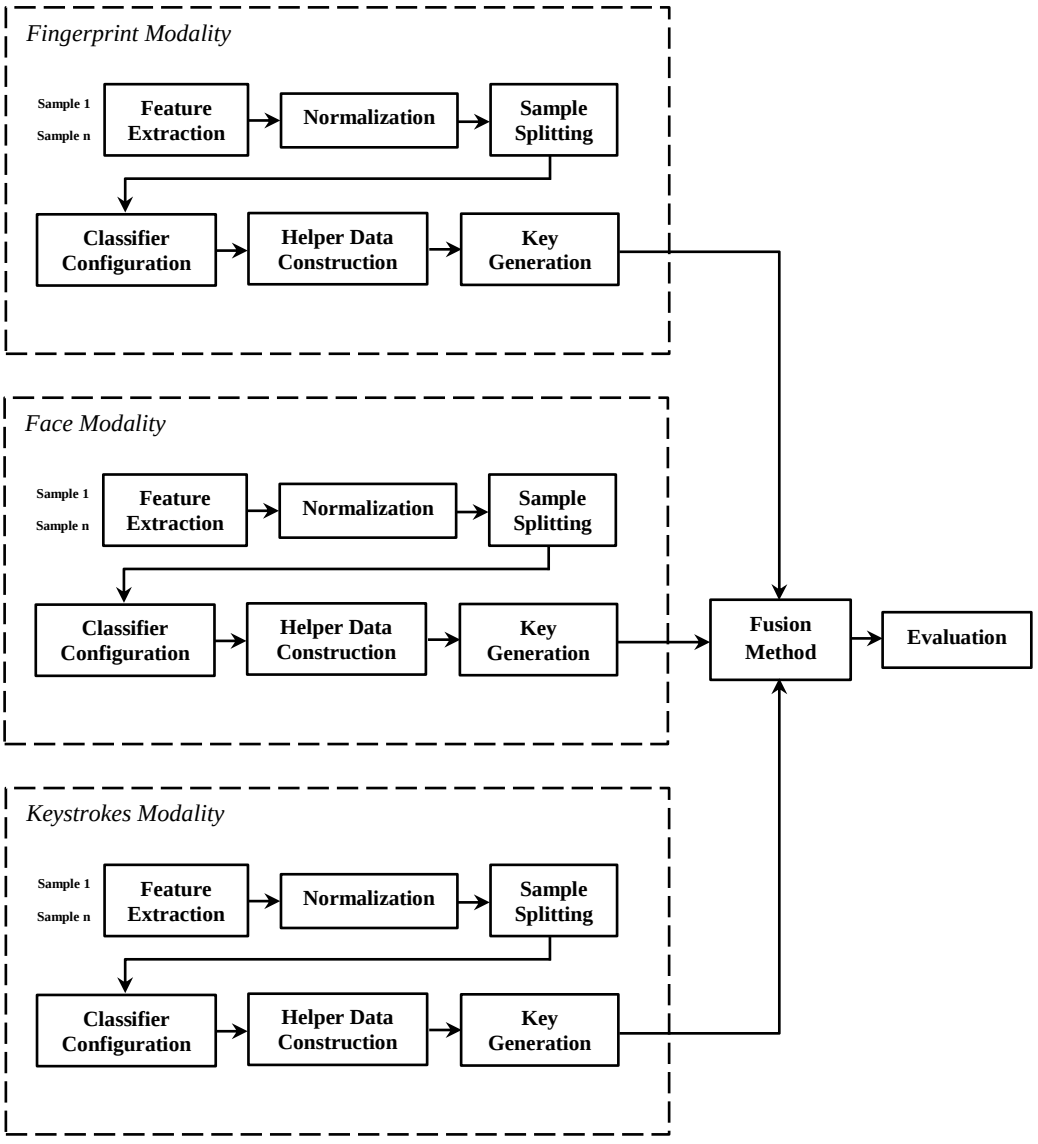


Figure 5. 1 Matching Fusion Process of Selective Biometric Modalities Using Majority Voting

5.3 Results and Analysis

Having implemented the presented investigations into multibiometric fusion approaches, the performance evaluation is presented in this section. The experiments were implemented and accomplished by writing and generating a set of python programming scripts. The following subsections are devoted to interpreting and analysing each experiment individually as follows:

5.3.1 Experiment 1: Transparent Multibiometric Key Generation by Feature-Level Fusion

The results and analysis of this experiment reveal the reliability of generating the bio-crypto key material by combing the selective biometric modalities using feature-level fusion. On the whole, the experimental results of this investigation demonstrate that the multibiometric approach at the feature stage has the capacity to create a more reliable and consistent bio-crypto key on a timely basis.

Table 5.4 shows the performance of the key generation by fusing all biometric modalities according to the sample unification procedures as follows:

Table 5. 4 Key Creation Performance via Combining All Biometric Approaches at Feature Level

Modalities	Sample Unification Procedure	FAR	FRR
Fingerprint + Face + Keystrokes	8-Sample	0.09%	61%
Fingerprint + Face + Keystrokes	50-Sample	0.02%	34.48%

Generally, it is evident from the tabulated results that both sample unification procedures amongst biometric modalities can generate a reliable and reproducible cryptographic key using the feature level fusion. The overall FAR rate indicates that

limited forgery attempts were taken place, and simultaneously the average FRR value is within the acceptable FRR figure of this research – less than 83%. On the other hand, incorporating the biometric modalities using the 50-sample unification procedure superiorly improves the key generation process. In comparison with the key generation accuracy on the basis of 8-sample unification, the FRR rate reduces by half, and the FAR value decreases by seven times. This can be explicated due to the limitation of incorporating few numbers of samples amongst the biometric modalities (only 8 samples).

Table 5.5 shows the performance of the biometric key generation by fusing a number of transparent biometric modalities reliant upon their availability over time (i.e. the key generation accuracy via combing all permutations of biometric approaches) as below:

Table 5. 5 Key Generation Performance by Incorporating Different Permutations of Biometric Modalities Using Feature Level Fusion

Modalities	FAR	FRR
Fingerprint + Face + Keystrokes	0.02%	34.48%
Fingerprint + Face	0.01%	67.69%
Fingerprint + Keystrokes	0.01%	67.02%
Face + Keystrokes	0.004%	64.29%
Fingerprint	0.94%	59.56%
Face	0.06%	62.69%
Keystrokes	0.06%	31.13%

As shown in Table 5.5, the experimental results describe that the combination of all three transparent biometric techniques outperforms the single biometric approaches in creating a bio-crypto key of 256-bit length. In accordance with this, whilst the FAR rate diminishes by four times versus the FAR rate of the individual approaches of face and keystrokes, it minimizes by seven times versus the FAR rate of the single

fingerprint technique. Simultaneously, the FRR figure of 34.48% on average confirms that all users had the capability to generate the bio-crypto key – evidently the FRR rate nearly decreases by half in comparison with the single modalities of fingerprint and face.

With regard to a combination of two biometric techniques, the empirical results overall demonstrate that a lower FAR rate is achieved in comparison with the FAR rates of all other permutations (i.e. single modality and three-biometric fusion). With 0.008% FAR rate on average across all two-biometric combinations, the minimal illegitimacy access was occurred. On the other hand, it would be expected that using a combination of two biometric modalities would outperform a single biometric approach in producing the bio-crypto key, but this was not the case. The FRR rate is rather higher than other single biometric techniques. This can be because one or two of the biometric feature vectors is very noisy resulting ultimately in producing the wrong bio-crypto key. Therefore, capturing the two quite consistent biometric samples would be ideal to increase the key generation accuracy. As shown in 4.2, this perspective would be achieved within the proposed approach as numerous samples can be transparently collected and the bad samples can be ruled out based on the key verification approach using a strong hash function. Despite this, the entire population succeeded to create the bio-cryptographic key of a 256-bit size as the FRR values are less than the tolerable figure of 83%. Using a unimodal biometric in particular the keystrokes would be insufficient for security purposes. Even if the FRR rates of combining two biometric techniques slightly increases without hindering the generation of a timely correct key, the FAR rates demonstrate that the encryption/decryption would be undertaken in a more secure manner.

It is also worth noting that there is no significant difference in the FRR rates across all combinations of two biometric approaches. A possible explanation is that the duplication of fingerprint samples in addition to ruling out a number of face and keystrokes samples (i.e. 33 facial samples and 10 keystroke samples) might be leading to the fixity in the FRR results.

Figure 5.2 comperes the individual FAR and FRR rates for each user in generating a bio-cryptographic key of 256-bit by combining all biometric modalities on the basis of 50-sample unification as follows:

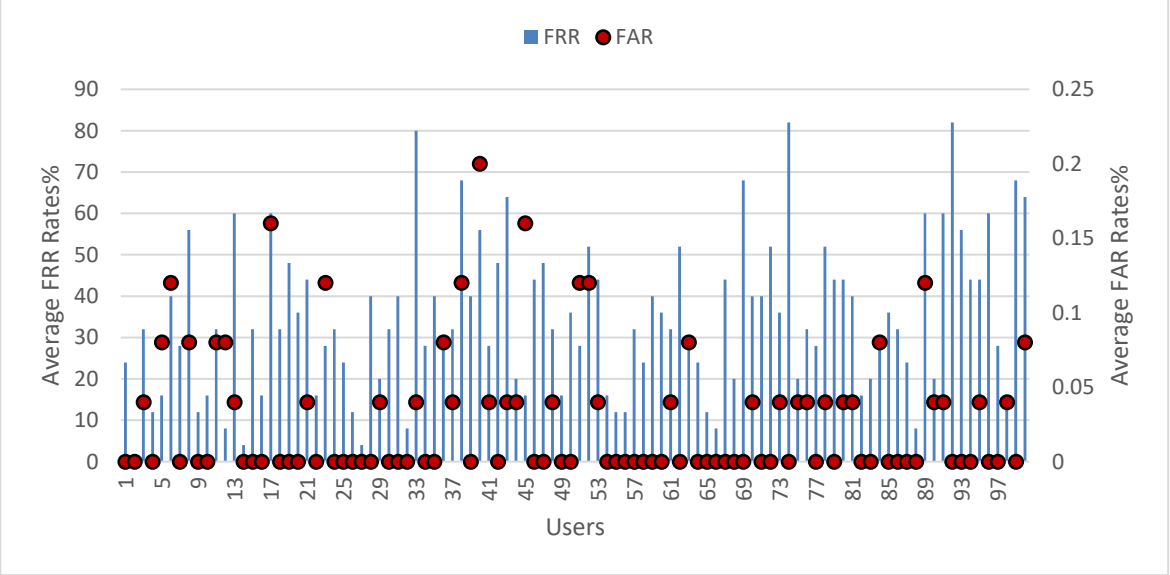


Figure 5. 2 Feature Fusion Performance of Three-Biometric Modalities via 50-Sample Unification for All Users

According to the chart, the significant majority of participants (80 users) had the ability to create the successful bio-crypto key using more than half of their samples with 27% FRR on average. At the same time, with a minimal FAR rate ranging from 0%-0.16%, 1962 samples of imposters failed to compromise the valid secret key (the whole samples set of the illegitimacy class is 79 imposters × 25 samples = 1975 samples). This indicates that the key generation reliability can be highly improved by

incorporating multiple biometric approaches. On the other hand, just under a quarter of population (20 users) utilized less than half of the biometric samples to produce the correct secret key – the FRR rate are 63% on average. At the same time, the FRR rate of 0.05 shows that only one sample of the invalid class samples which were 19 imposters ^{x} 25 samples = 475 samples can hack the correct key. However, regardless how many samples were effective in generating the cryptographic key, the multibiometric via feature-level fusion demonstrates that all users succeeded to produce the key of a 256-bit length.

5.3.2 Experiment 2: Transparent Multibiometric Key Generation by Matching-Level Fusion

The empirical results and analysis of this experiment determine the robustness of producing a biometric key of 256-bit via integrating the contributory biometric approaches using matching-level fusion. Generally, the experimental outcomes of this investigation confirm that some users yet cannot generate a transparent secret key on time using the multibiometric approach at the matching phase.

Table 5.6 reveals the effectiveness of generating a key by combining all biometric techniques according to the sample unification procedure as below:

Table 5. 6 Key Generation Performance by Incorporating All Biometric Approaches at Matching Level

Modalities	Sample Unification Procedure	FAR	FRR
Fingerprint + Face + Keystrokes	8-Sample	0%	97.4%
Fingerprint + Face + Keystrokes	50-Sample	0%	81.91%

Based on the FAR and FRR figures, the above-tabulated results describe that the matching approach utilising the 50-sample unification procedure is superior in

creating a bio-crypto key of a 256-bit length. With the multibiometric fusion being applied on the basis of 50-sample unification, the FRR rate indicates that the correct keys were produced to the significant genuine users (80 users), and concurrently the FAR rate reveals that the wrong keys were produced to all imposters. On the other hand, combining the biometric modalities on the basis of the 8-sample unification negatively affect the key generation process. Just under a quarter of the whole population (13 users) had the ability to encrypt/decrypt data in a non-intrusive and continuous fashion; however, the vast majority of participants did not succeed to do so on a timely basis. This might be because of unifying poor quality of samples in between the fingerprint, face and keystrokes techniques in addition to the limitation of consolidating few samples (i.e. 8 samples from each biometric). In addition, 75 and 52 facial and keystrokes samples were respectively excluded which might be the effective ones in generating the desired bio-crypto key of 256-bit length.

Table 5.7 compares the performance of the biometric key generation by fusing a number of transparent biometric modalities reliant upon their availability over time (i.e. the key generation accuracy via combing all permutations of biometric approaches) as follows:

Table 5. 7 Key Creation Effectiveness through Combining Different Permutations of Biometric Modalities Using Matching Level Fusion

Modalities	FAR	FRR
Fingerprint + Face + Keystrokes	0%	81.91%
Fingerprint + Face	0.002%	77.37%
Fingerprint + Keystrokes	0.0008%	74.71%
Face + Keystrokes	0.0047%	67.98%
Fingerprint	0.94%	59.56%
Face	0.06%	62.69%
Keystrokes	0.06%	31.13%

It is clear from the results that the incorporation of all selective biometric modalities do appear to be less effective in generating the key of a 256-bit amongst all permutations of biometric approaches, with 81.91% FRR rate on average. A possible explication is that balancing the legitimate and illegitimate classes based on the replication of only 4 fingerprint samples could have a negative impact upon the key generation process. Another possible reason is that the 33 face samples and 10 keystrokes samples which are excluded might be the successful ones in generating the correct key under which the FRR rate could be fallen down. On the other hand, the multibiometric approach of all three modalities accomplishes the superior performance in overcoming the illegitimate key generation by forgers among all biometric permutations. Evidently, the FAR rate of 0% demonstrates that no imposter can hack the cryptographic key.

In terms of two-biometric fusion, the key generation performance is pretty encouraging in comparison with the single biometric. The FAR rates reveal that the minimal forgery attempts were taken place. At the same time, whilst a far lower proportion of the entire population (2 users only) cannot generate the correct key, 11 individuals failed to create the same valid key from the single fingerprint and face techniques. Although the FRR rates are rather higher, the number of users that were able to generate the bio-crypto key is lower. This could be interpreted due to the fact that one of the biometric modalities gave a negative accuracy and degraded the incorporation of two biometric techniques. Therefore, ensuring that the effective biometric samples from both modalities are concurrently presented to the multibiometric approach is quite crucial to escalate the key generation accuracy.

Figure 5.3 comperes the individual FAR and FRR rates for each user in producing a bio-cryptographic key of 256-bit by incorporating all biometric modalities on the basis of 50-sample unification as below:

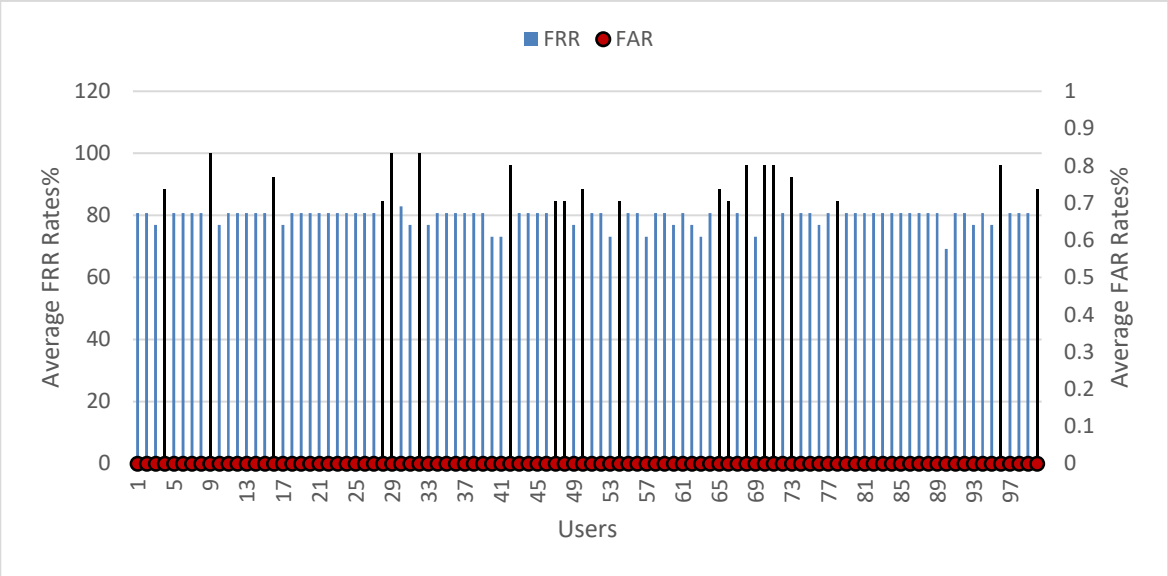


Figure 5. 3 Matching Fusion Performance of Three-Biometric Modalities via 50-Sample Unification for All Users

It is clear from the chart that a far greater proportion of participants (80 users) can have the capacity to transparently and repeatedly generate a bio-cryptographic key of a 256-bit length. Simultaneously, with 0% FAR rate on average, no imposter succeeded in forging the correct key with the purpose of cracking the encrypted data. In contrast, just under a quarter of the entire population (20 users) failed to create the bio-crypto key. This might be because of excluding 33 and 10 biometric samples approximately from the face and keystrokes modalities respectively.

5.4 Discussion

Holistically, the research findings show that the combination of the fingerprint, face, and keystrokes techniques all together outperform the individual biometric in creating

the bio-crypto key of a 256-bit to cipher/decipher data in a realistic approach. Evidently, the experimental results reveal that the stored data of all 100 users within cloud storage would be protected in a more reliable way without any inconveniences. This means that the key generation accuracy has been positively escalated. At the same time, the security aspects in terms of combating the forgery attacks and features guessing attacks would be surely underpinned.

With regards to the multibiometric performance, the feature level fusion overall outperforms the matching level fusion at producing the valid correct key with limited illegitimacy attempts. By incorporating the three biometric modalities on the basis of 50-sample unification, the FRR rate of the multibiometric approach via feature level fusion decreases 47 times approximately in comparison with the FRR rate of doing so by the matching level fusion. In addition to this, whilst the entire population (100 individuals) had the ability to create the bio-cryptographic key through the biometric combination at the feature phase, 20 users did not succeed to do so at the matching stage. In terms of the biometric incorporation reliant upon the 8-sample unification procedure, the feature level fusion has been also more superior to generate the bio-crypto key than the matching level fusion. In accordance with this, all users succeed to create the correct key at the feature stage; however, the whole population failed to establish the same valid key at the matching point.

Regarding the two-biometric fusion, the empirical results demonstrate that the feature-level fusion accomplishes the superiority in creating the secret key of a 256-bit in comparison with the matching-level fusion. While all 100 participants can generate the cryptographic key by combining any two biometric modalities at the

feature phase with approximately 66% FRR rate on average, 98 users can do so at the matching stage with nearly 72% FRR rate on average.

From the security perspective, the multibiometric approach on the whole whether three-biometric fusion or two-biometric fusion reduces illegitimacy instances in compromising the valid bio-crypto key. With three-biometric approach being applied at the feature level on the basis of 50-sample unification, the average FAR rate was 0.02%. On the other hand, by combining all the three biometric techniques at the matching level, the FAR rate was 0% on average using the same sample unification procedure. As such, the FAR rate of the multibiometric key generation approach evidently outperforms the single biometric modalities in which the average FAR rate was 0.94%, 0.06%, and 0.06% for fingerprint, face, and keystrokes respectively. The multibiometric key generation approach would be able to resist the potential spoofing attacks. Furthermore, the multiple biometric incorporation underpins the entropy factor where more than one feature vector is employed to generate the bio-crypto key. Accordingly, the number of possible combinations from each particular feature vector is increased; thereby, the feature guessing attacks can be overcome. The effective entropy of a multibiometric approach is measured by determining the difference between the maximum and the minimum values a biometric feature can have. Then, the $\log_2$ for the product of those values is taken to evaluate the entropy for a biometric technique in an effective bitlength. Accordingly, the entropy values from each biometric approach are multiplied to evaluate the entropy of a multibiometric key generation approach. Table 5.8 shows the effective entropy of biometric features of all approaches (i.e. single biometric, two-biometric, and three-biometric).

Table 5. 8 Effective Entropy (Bitlength) of Single, Two, and Three Biometric Approaches

Approach	Effective Entropy
Fingerprint + Face + Keystrokes	6567438592
Fingerprint + Face	4366648
Fingerprint + Keystrokes	5740768
Face + Keystrokes	1720576
Fingerprint	3817
Face	1144
Keystrokes	1504

5.5 Conclusion

This chapter has demonstrated through the experimentation the outperformance of generating the bio-crypto key of a 256-bit from the transparent multibiometric approach. Using the multibiometric encryption approach would increase the reliability of cloud storage technology where more usable and secure framework will be enabled with no doubt. In general, the empirical outcomes show that the multibiometric key generation using the feature level fusion is better than the matching level fusion in generating the cryptographic key. In addition, the experiments revealed that the deficiency of biometric data has a huge impact upon the key generation performance; especially within the multibiometric approach via the matching level fusion. Although there was a limitation regarding experimenting few biometric samples, numerous samples would be spontaneously collected in reality leading to the opportunity of enhancing the key generation performance.

Chapter Six: Transparent Bio-Cryptosystem Architecture

6.1 Introduction

Having successfully demonstrated the empirical foundation for generating a robust non-intrusive cryptographic key via transparent biometric approaches, this chapter is devoted to proposing and designing an innovative bio-cryptosystem framework. Whilst the bio-cryptosystem presented in this chapter has applicability to any stored data, it is particularly relevant for cloud-based storage given the lack of additional security controls in place. In order to lay out the architectural framework, a number of fundamental requirements would be identified – with specific issues in mind regarding security, usability, scalability and efficiency aiming at developing a well-considered platform. As a result, appropriate components, processes and mechanisms would be employed to manage and maintain a reliable key generation solution and to ultimately provide a seamless intelligent encryption. In addition, a number of operational aspects that a real-time system would require will be also taken into consideration and accordingly discussed in an analytical view for approaching an effective scenario.

An innovative cryptographic technology would be non-intrusively undertaken by generating a bio-crypto key using the stored helper data on enrolment and the live biometric features. In view of storing local helper data, there would be an issue in accessing the online storage service from multiple devices – as the local helper data would be only stored on a single device. With the promising characteristics being offered by cloud computing in terms of universality, connectivity, scalability and flexibility, it would be possible to tackle the issue of the helper data storage. As such,

this public data can be stored centrally within the Cloud to provide a universal access for all users' devices and then undertake the encryption service. The proposed bio-cryptosystem architecture presents an advanced secret key generator – that is capable to create a key on the fly without storing it anywhere in a transparent fashion. Nonetheless, using such an innovative system will also introduce issues ranging from security, privacy to automation that have to be addressed and overcome.

The subsequent sections start with identifying the system architecture requirements dependant upon the analysis and the experimental outcomes from the previous chapters. Then, a comprehensive explanation of the system components, processes and mechanisms is presented - focusing upon the security and usability aspects with a view to ensuring a robust and convenient cryptographic framework. A number of operational considerations are also addressed and conceptually resolved with the purpose of reinforcing the system operation in reality.

6.2 System Requirements

Prior to the architectural design, a number of system requirements should be specified in order to offer an effective solution. Based upon the critical analysis of the literature review presented in Chapter 3 in addition to the results of the experimentation phase conducted in Chapters 4 and 5, the transparent bio-cryptosystem requirements have been identified by the following:

1- Advanced measures of security management

   Secure management procedures are needed with the aim of lowering the sensitivity of the helper/public data that has to be stored at some location to facilitate the key regeneration for encryption or decryption. Of course, the

storage of the public data should not help imposters to reveal any information to breach the system. As shown within the experimental phase (Chapters 4 and 5), the methodological approach exploits the conception of biometric recognition by driving the multilabel classification problem to cope with this requirement in a secure manner. As such, a pattern classification technique (i.e. a neural network algorithm) is taken advantage to create the cryptographic key without storing sensitive data. From a different perspective, it is also evident from the experiments of Chapter 4 that the dependence upon one biometric approach can be deemed as an unsatisfactory in terms of security in addition to the performance (which is discussed in requirement 3). Single biometric modality can be defeated easily because of low individuality, high forgery attempts and lack of universality. For example, face biometric system is vulnerable to be hacked by spoof attacks. Another instance is that using the keystroke dynamics technique alone is fairly fragile. As discussed in 2.6.1.7, the distinctive actions of keystrokes seem to be insufficient for verifying a person, and they are an inadequate to be applied for security purposes. As demonstrated from the experiments of Chapter 5, the multibiometrics fusion can arguably contribute to overcome or at least alleviate these shortcomings. From another viewpoint, in a bio-cryptosystem under which the feature vector in particular acts (directly or indirectly) as an encryption enabler, the capacity of the approach to overcoming brute force attack is essential. The biometric entropy is a measure of the number of possible combinations a particular feature vector can have. It can reflect an indication in determining the effort required to brute force a biometric feature vector by an attacker. As such, the biometric entropy reinforcement is a very

important requirement that should be taken into account when designing a bio-cryptographic system. This requirement can be achieved by incorporating significant biometric features to reinforce the biometric entropy. The multibiometric cryptographic approach in Chapter 5 has accomplished a robust biometric entropy (bitlength) reaches up to 6567438592.

2- High level of transparent and continuous operation

Relying upon third-party cryptographic applications to provide an additional user-oriented level of protection can be arguably cumbersome. The cumbersomeness issue is posed when each file needs to be encrypted/decrypted manually in addition to administering many keys. As a consequence, there is apparently a need for a transparent, continuous and convenient approach for generating a robust cryptographic key to accomplish seamless encryption. Accordingly, the concepts of biometric cryptography and transparent biometrics are incorporated with one another to remove the usability issues in terms of having to present biometric credentials each time a file needs to be encrypted/decrypted or recalling long complex keys. Therefore, the application of transparent biometric cryptography would enable the generation of a non-intrusive timely cryptographic key and prevent the person from having to present cryptographic credentials (e.g. secret keys, passwords and biometrics).

3- Acceptable degree of performance

The system capacity for generating a valid reproducible cryptographic key to the genuine person only is a crucial requirement. Referring back to the requirement 1, using one biometric technique will be obviously unreliable. In addition to the downsides of the single biometric modality with regard to

security, it can simultaneously impact the system effectiveness in a negative manner owing to poor uniqueness, high error rate and lack of universality. For instance, face biometric modality is affected by position, expressions and the amount of present illumination; thus, the face biometric system can be easily defeated. According to the experiments of Chapter 5, the multiple biometric integration would certainly improve the key generation performance. On the whole, the key-findings from the empirical results (Chapter 5) demonstrate that a valid repeatable cryptographic key to the legitimate person of 256-bit length can be optimistically generated to encrypt/decrypt the stored data in reality.

4- Diverse key generation approach

This requirement means that various cryptographic keys have to be produced for encrypting or decrypting each document within the cloud storage paradigm. The key diversity requirement is fairly important characteristic by which the confidentiality and privacy aspects are boosted – thus if one document is hacked, the other one would not be influenced as it is encrypted by a different key. The random key generation approaches can handle this requirement to ideally create numerous keys for each document. On the other hand, additional security mechanisms should be given in place in order to provide an acceptable protection to the assets of a random key generation approach.

5- Robust revocation procedure

The revocability refers to the capability of cancelling the generated key in case of hacking and reissue another one. Referring to the previous requirement, due to the generation of diverse cryptographic keys, a comprehensive and tactical revocation procedure should be considered with a view to revoking

any key in case of compromise. Fortunately, the bio-cryptographic domain supports the key revocability characteristic where neither the biometric data nor the key will be stored at some location. Instead of that, public data are stored to aid in generating a timely cryptographic key thus facilitating to accomplish the revocation requirement within the proposed system architecture.

6- **Universal and interoperable solution for cross-platform usage**

Given the necessity to access the cloud storage service from multiple devices, universal and interoperable solution is required to undertake the encryption technology on different personal devises. The existing cloud paradigm already supports universality and interoperability characteristics that allow wide range of technological devices (e.g. desktop computer, laptop, tablet and/or mobile phone) to be connected to the storage service. Accordingly, this requirement is very important to achieve; otherwise, the system architecture will be incomplete from a usability perspective – such deficiency could badly impact the system acceptability and adoptability. The cloud commuting can be introduced as a solution with a view to fulfilling this requirement. Thus, the entire core techniques that are applied to facilitate in generating the bio-crypto key would be placed within the Cloud to achieve transparent encryption. Another solution can be used by conducting a registration session upon each device. That is, while signing up, there would be collector agents to capture the biometrical signals transparently once for enrolment. Then, a user profile would be used to construct helper data and create the desired key on a timely basis in the meantime.

6.3 System Architecture

Having identified the system requirements and characteristics, the bio-cryptosystem architecture has been proposed accordingly with the aim of maximizing the security and usability aspects. The innovative architecture system aims to offer a convenient user-oriented cryptographic framework to additionally secure the data privacy and confidentiality within cloud storage. The core approach of the proposed system overcomes the poor and cumbersome secret keys and eradicates the inconvenience issues of the third-party cryptographic tools – via applying the transparent biometrics modalities within biometric cystography. Thereby, the cloud storage subscribers would be in charge of protecting their data in a more usable fashion – where a transparent repeatable bio-crypto key would be generated for encryption/decryption. At the same time, this key would be reliably and consistently generated on the fly via employing public data without storing it at some location to reinforce the security factor.

The bio-cryptosystem architecture implements a multibiometric topology using a variety of transparent biometric modalities, such as fingerprint, face (physiological) and keystroke dynamics (behavioural). The framework is devised reliant upon a hybrid approach – using an integration of multi-modality and multi-sample sources. In particular by applying multiple transparent biometric techniques, the proposed system architecture would be capable to robustly perform encryption or decryption process in a seamless fashion. In addition to this, the presented bio-cryptosystem can be highly modular – where a wide range of physiological and/or behavioural biometric approaches can be also applied as appropriate with a view to elevating the system security and accuracy.

Figure 6.1 depicts the architectural bio-cryptosystem which fundamentally includes processing engines and agents situated within the parties of the client and the provider.

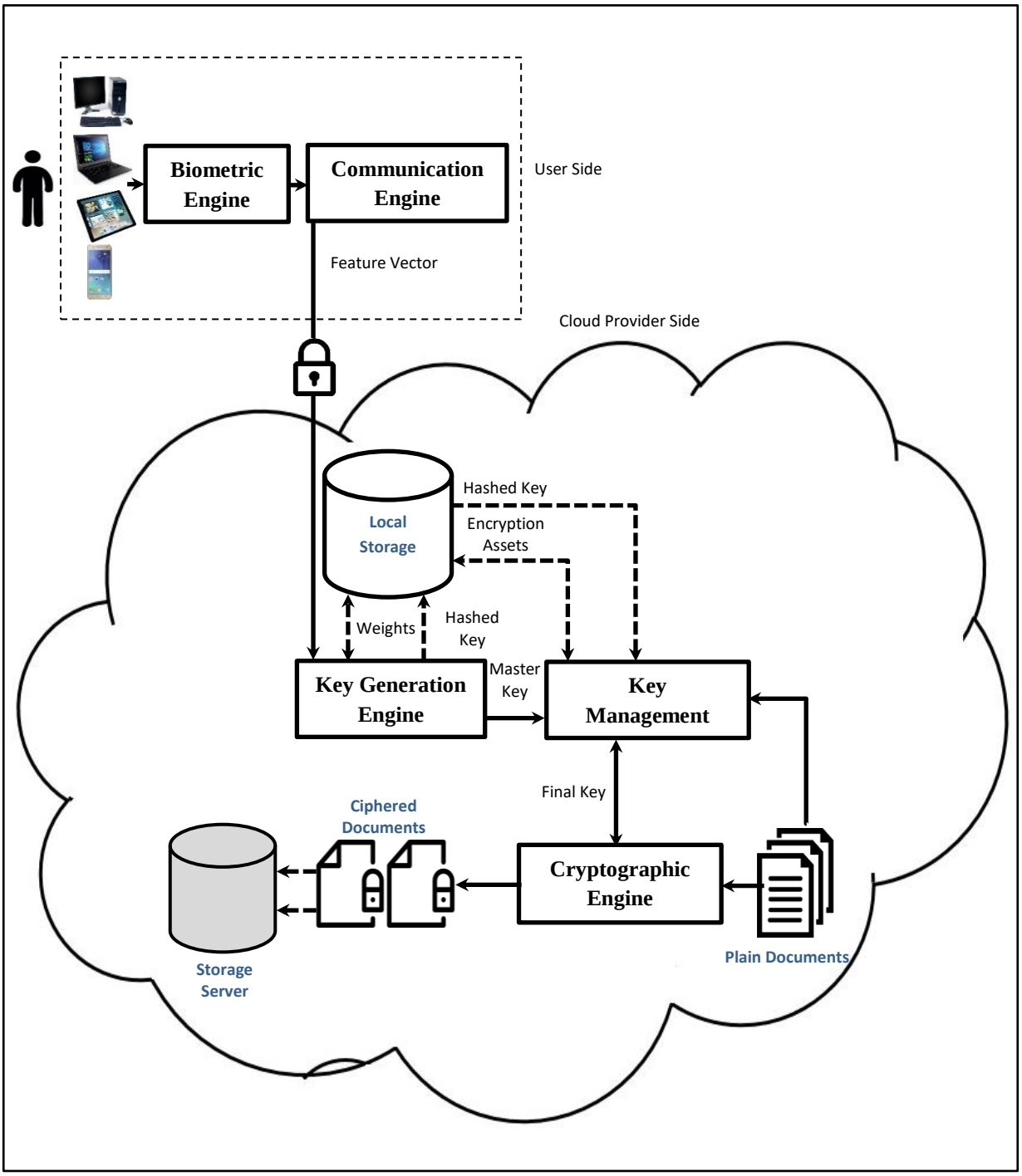


Figure 6. 1 An Innovative Model of Multi-Biometric Cryptography Undertaken at the Cloud Side

In order to understand the presented architectural system, the scenario of the cloud data protection is described. At first, the authentic cloud user should log in to the service via the standard verification protocol. Authentication is the frontline countermeasure of ensuring that only the genuine user is granted access into the cloud storage, and accordingly a robust authentication technology should be provided in order to accomplish a confidential cryptographic framework. Given encouraging outcomes from transparent biometric verification in terms of accuracy and usability being demonstrated, it can be combined with two-factor authentication with a view to ensure a trusted, secure and reliable access. In the meantime, while the cloud user undertakes particular cloud storage service activities, his/her biometric data, which does not require the explicit interaction with the system, are collected in a non-intrusive manner. Cloud-based storage activities can be Add, Delete, Edit, Download, Update, Rename, Read, Write, or any other activity that could be undertaken by the cloud subscriber. The biometric modalities are acquired via the Biometric Collector Agent and then directed into a number of input sampling channels. Following this, a number of the feature extraction techniques are applied to generate the optimal feature vector – this stage represents the Feature Extractor Agent. Both the Biometric Collector Agent and the Feature Extractor Agent belong to the **Biometric Engine**. Subsequently, the **Communication Engine** sends over the feature vector into the cloud provider which undertakes all remaining processing and responsibilities with the aim of setting out the key generation process for multiple users' devices.

The cloud provider party, being the backbone of the transparent biometric key generation process, fundamentally consists of three engines: **Key Generation**

Engine, **Key Management Engine** and **Cryptographic Engine**. The **Key Generation Engine** includes the classification approach which is used to generate/identify the bio-crypto key (i.e. master key) reliant upon the live features and the training parameters (i.e. weights). The **Key Management Engine** is responsible for managing and maintaining the master key – leading ultimately to the launch of numerous valid bio-crypto keys on the basis of successful key verification. In the **Cryptographic Engine**, each valid bio-crypto key is used to seamlessly encrypt/decrypt each document within the cloud storage in a non-invasive way. Eventually the cloud storage provider will be responsible for storing the secured document.

In view of potential impacts upon the previous innovative system architecture (Figure 6.1) in terms of security and scalability, another bio-cryptosystem architecture for cloud-based storage is presented. The use of the **Communication Engine** to transmit the biometrical signals to a Trusted Third Party (TTP) with a view to undertaking a secure transit to the cloud provider might raise security issues related with trust. In addition, storing helper data for each user subscribed with the cloud storage service could pose a burden upon the cloud provider – resulting in negatively affecting the scalability characteristic. In order to cope with the above-mentioned issues, another system architecture is designed concentrating upon storing the helper data, which would aid in generating the bio-crypto key, at the client side. Such a bio-cryptosystem would eradicate the need of using a communication engine that hands in the biometric features from the user side to the cloud provider through a Trusted Third Party (TTP).

Figure 6.2 shows the architectural bio-cryptosystem that encompasses processing components located only at the client party as follows:

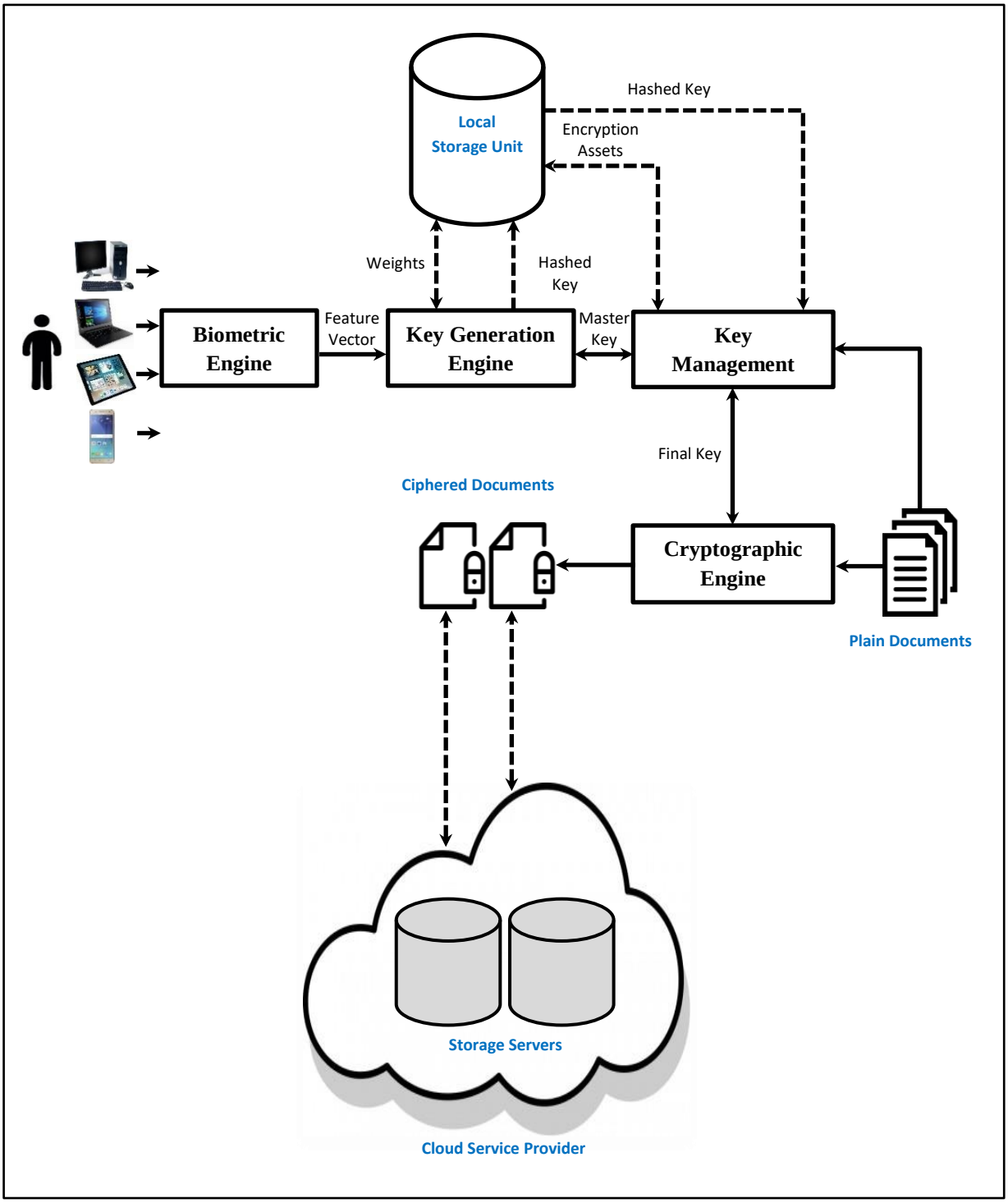


Figure 6. 2 An Innovative Model of Multi-Biometric Cryptography Undertaken at the User Side

As illustrated in Chapter 1, the local storage units can be regarded more secure than cloud storage - where the former employs physical and logical security controls before granting access to the stored data. As such, the framework of the above bio-cryptosystem offers the local storage solution with the aim of lowering security vulnerabilities (i.e. the exploitation of personal storage units, such as hard drives) whereas the existing cloud paradigm lacks the potential afforded to local storage strategies. Figure 6.2 describes that the client party, being the heart of the transparent key generation process, comprises all the previous processing engines except the **Communication Engine**. These are the **Biometric Engine**, **Key Generation Engine**, **Key Management Engine** and **Cryptographic Engine**. As the biometrical signals would be treated at the user side, there is no actual need to utilize the **Communication Engine**. The tasks of these processing engines are as same as in the former presented architecture. In the latter proposed system architecture, the user would sign up an account on cloud storage, for instance, from his laptop. Then, the transparent encryption would be normally undertaker on the client party as in the former architecture without transmitting the biometrical singles to the cloud provider. However, when the subscriber wants to log in to the service from another device such as a mobile phone, he/she must either submit the biometric data to the system or the system should capture the biomedical data transparently during an enrolment session for only once - as same as in Apple iPhone. It is worth noting that the Apple iPhone user can turn into using the fingerprint biometric as a passcode to unlock the device (see section 2.6.1.1). Having collected biometric samples on enrolment, they would be utilized for building another user profile and training the classifier to construct an indispensable public/helper data which will be stored on the mobile phone this time. This data would be used alongside with the fresh biometric samples

to generate the bio-crypto key on encryption/decryption. The same scenario would occur with other new devices. Although the second architecture tackles the security and scalability issues, it yet presents a potential security concern in terms of storing multiple helper data upon different personal devices. Given the cons and pros from both architectures, one of them would be adopted depending upon the user desirability and the provider initiatives.

The following section explains in detail the architectural system components in order to fulfil its requirements thereby maintaining security and minimising inconvenience.

6.4 System Architecture Components

The presented architectural systems consist of common processing engines cooperating to ultimately enable more usable and secure cryptography, albeit the first system architecture included another component – the **Communication Engine**. With the framework of each proposed system being taken into account, the processing engines are explained by the following:

1. **Non-Intrusive Biometric Engine**

   The Non-Intrusive Biometric Engine basically comprises two agents: the **Collector Agent** and the **Feature Extractor Agent**. The primary task of the Collector Agent is to detect and collect the biometric information of a user both physiological and behavioural (e.g. fingerprint, face, voice, keystrokes, and behavioural profiling). However, it is not possible to ensure that all of the biometric modalities will be always acquired because biometric samples can be only captured if they are available in a non-intrusive manner. Despite this, the Collector Agent will highly likely collect some biometric data as long as the

user would interact with the cloud storage account through reading, writing, or editing files. In this context, the system would have the possible biometric samples to generate the bio-crypto key. On the other hand, the Feature Extractor Agent is responsible for pre-processing the collective biometric data and extracting the biometric features. Figure 6.3 depicts the non-intrusive biometric engine bloke diagram as follows:

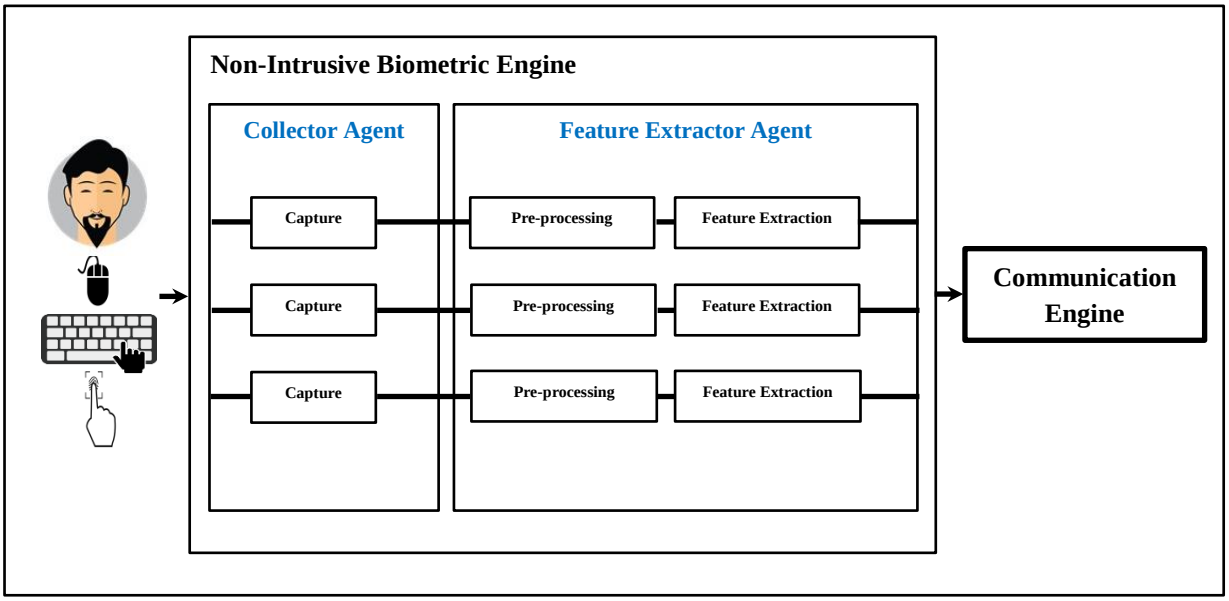


Figure 6. 3 Non-Intrusive Biometric Engine

The bloke diagram of the biometric engine illustrates that there are a number of agents within the collector agent situated to acquire the samples of biometric modalities. Each agent continuously captures its input samples in the background of a device in a non-intrusive manner. Once the agents have accomplished the capturing of the applied biometric samples, the Collector Agent then proceeds to the forth agent. That is, the Feature Extractor Agent pre-processes the biometric data and accordingly extracts the more reliable biometric features. Reliant upon the individual biometric technique, the pre-processing stage and further processing steps will be performed differently.

For fingerprint and face, a number of analysis, enhancement, and alignment approaches can be used to ameliorate the quality of the biometric image samples and to transform them into a common version. On the other hand, the pre-processing task of input data obtained by the keystroke dynamics technique involves calculating the duration time, performing outlier removal and normalisation of the timing vector. The feature extraction process then extracts the biometric features from the processed samples and converts them into feature vectors.

The literature review conducted in Chapter 3 indicates that significant amount of research tends to remove numerous biometric features with a view to overcoming the biometric variations. Whilst this perspective could improve the biometric key generation performance, it can badly affect the biometric entropy leading to feature guessing attack. As a consequence, incorporating numerous independent biometric features would present many combination values from a feature vector which cannot be guessed by an attacker in the meantime. As such, supporting the biometric entropy is a crucial requirement that should be fulfilled. The experimental phase within Chapter 4 and 5 demonstrates an effective biometric entropy by applying a number of commercial feature extraction techniques.

On enrolment, the extracted source feature vectors are used for building a user profile and training the classifier to construct an indispensable public/helper data. However, the extracted fresh feature vectors on encryption are utilized to generate a constant bio-crypto key on a timely basis in the meantime. There is no doubt that the performance of a biometric key generation system is directly related to the quality of the biometric samples,

and this could be positively handled by the pre-processing phase. However, it would be very efficient to check the quality of the sample within the Collector Agent prior to undertaking the pre-processing. The Collector Agent can have the capacity to check the quality of the captured sample in advance by seeking a proper procedure. For instance, The Collector Agent for the fingerprint and face can utilize predefined image resolution threshold to check the quality of the collective samples. Furthermore, the Collector Agent for the keystroke analysis modality could be capable to check the quality of the captured sample by determining the event duration thus indicating if it is normal or higher than normal, for instance. This illustrates that the user might be interrupted while typing data. Nevertheless, the issue of how to measure the quality of the biometric samples is the topic of considerable discussion and research within the design of biometric modalities, rather than a wider architectural issue.

From the multibiometric perspective, the system architecture is not confined to the selective biometric approaches in this research only; thus, it can adapt any new potential biometric modalities that can be acquired while using cloud storage. Accordingly, the Biometric Engine would be flexibly and adaptably mechanized seeking to incorporate other transparent biometric techniques. To this end, using biometric standards such as ISO (Standardization, 2005) is required. This means that other existing or emerging transparent biometric approaches can be applied within the proposed system architecture as long as they compile with ISO standards (i.e. ISO 19794, 19785, 19784). What is more, with the possibility of having classification and encryption being undertaken by the cloud provider, a wide range of devices can use such a

compatible system built upon using biometric standards. That is, the cloud user would have the capability to log in to his/her account from any device. These devises vary in regard of their hardware configuration and operating system (e.g. desktop, laptop smartphone, tablet, and so on). Therefore, utilizing ISO biometric standards would allow to undertake a resilient, modular, and compatible framework.

2. Communication Engine

Given the potential of having encryption being undertaken by the cloud service provider, the **Communication Engine** will be utilized as a secure communication channel to communicate between the client and the cloud parties for an agreeable objective. That is, the user needs to trust the service provider based upon the available agreements and policies. The communication process occurs in a trusted framework to a Trusted Third Party (TTP) and via the Secure Sockets Layer (SSL) security technology. As such, whenever the cloud side encrypts/decrypts the users' files, the Communication Engine would be activated on that basis. The Communication Engine will specifically transmit the biometric features, which have been collected, pre-processed and extracted by using the Non-Intrusive Biometric Engine, into the cloud provider. The service provider will subsequently undertake the cryptographic framework in a secure and usable manner (see figure 6.1). It is worth noting that the Communication Engine will not be used if the encryption/decryption is undertaken at the client side.

3. Key Generation Engine

The essential functionality of the **Key Generation Engine** is to establish a master bio-crypto key for the legitimate user only. When the encryption

framework is undertaken at the cloud side, this engine will apply the biometric key generation process via receiving the pre-processed feature vectors being handled within the **Non-Intrusive Biometric Engine** and transmitted by the **Communication Engine**. On the other hand, if the encryption framework is undertaken at the user side, the biometric key generation process will create the master bio-crypto key on the user's device itself only (as presented in Figure 6.2). On the whole whether the encryption would be applied at the client party or at the cloud side, this engine has to have the capacity to deal with single or multiple biometric modalities. This would ensure that the key generation process can be performed even if all of the selective biometric modalities are unavailable. As such, the **Key Generation Engine** generally comprises several agents. These agents are fundamentally classified into Biometric Key Agent and Multibiometric Key Agent. The former (**Biometric Key Agent**) creates a bio-crypto key via single biometric approach; however, the latter (**Multibiometric Key Agent**) generates a biometric key by using multiple biometric approaches as depicted in Figure 6.4. With the contributing biometric modalities being applied in this study, the **Key Generation Engine** would be capable to totally operate 7 types of key agents. These types are explained by the following:

- Fingerprint biometric key agent
- Face biometric key agent
- Keystroke dynamics biometric key agent
- Multibiometric key agent based on fingerprint, face and keystroke dynamics
- Multibiometric key agent based on fingerprint and face

- Multibiometric key agent reliant on fingerprint and keystroke dynamics
- Multibiometric key agent depending on face and keystroke dynamics

As illustrated in Chapter 2, a source should be provided to combine some data and thereafter develop a multiple biometric system. Since many biometric samples can be collected in a transparent fashion from one or more than one biometric modality, the following sources are adopted on the core proposed system framework – where one or set of them would be implemented as appropriate:

- **Multi-Modality Source**: building a single sample of more than one modality to tackle the potential weaknesses of some biometric techniques or acquisition devices.
- **Multi-Sample Source**: building multiple inputs of the same modality to have a well-informed identity and to offset the existing samples of bad quality.
- **Hybrid Source**: dynamically building single or multiple samples from different modalities. This could probably fine-tune the approach in generating the desired bio-crypto key, crafting a more multi-layered method.

According to the discussion in Chapter 2 section 2.5, these samples have to be incorporated effectively at a certain phase (i.e. sensor, feature, matching score, and/or decision level) within the biometric system. In this context, the biometric fusion overall aims to reinforce the capacity of generating a timely repeatable bio-crypto key for a seamless encryption. The experiments conducted in Chapter 5 investigated two approaches of biometric fusion.

These are feature level-fusion and matching-level fusion. The former can have the capability to escalate the accuracy and security of the biometric key generation. In addition to this, the latter has the merit of encompassing any other biometric modalities/classifiers without the need to re-train the system from the scratch. The experimental results, however, demonstrate that the feature-level fusion is more robust than the matching-level fusion based on the experimental data. As such, it is recommended that the applied biometric fusion method within the design of the Multibiometric Key Agent would be the feature-level fusion, although such a research finding could not be generalized. This is because the biometric data in reality would be different, but at least there is a tangible and promising basis.

From another aspect, the **Multibiometric Key Agent** can possibly generate the correct key even with a potential number of rejected samples. Whist the rejected biometric samples might be presented by an imposter, constructing temporary helper data via re-training the suspicious samples would aid the system to have a good indication about the real interactions of the legitimate user resulting in ameliorating the key generation performance. Contrarily, those sample will be removed from the local storage unite if the overall accuracy had indicated that the samples belong to an adversary. Figure 6.4 shows the block diagram of the Key Generation Engine as follows:

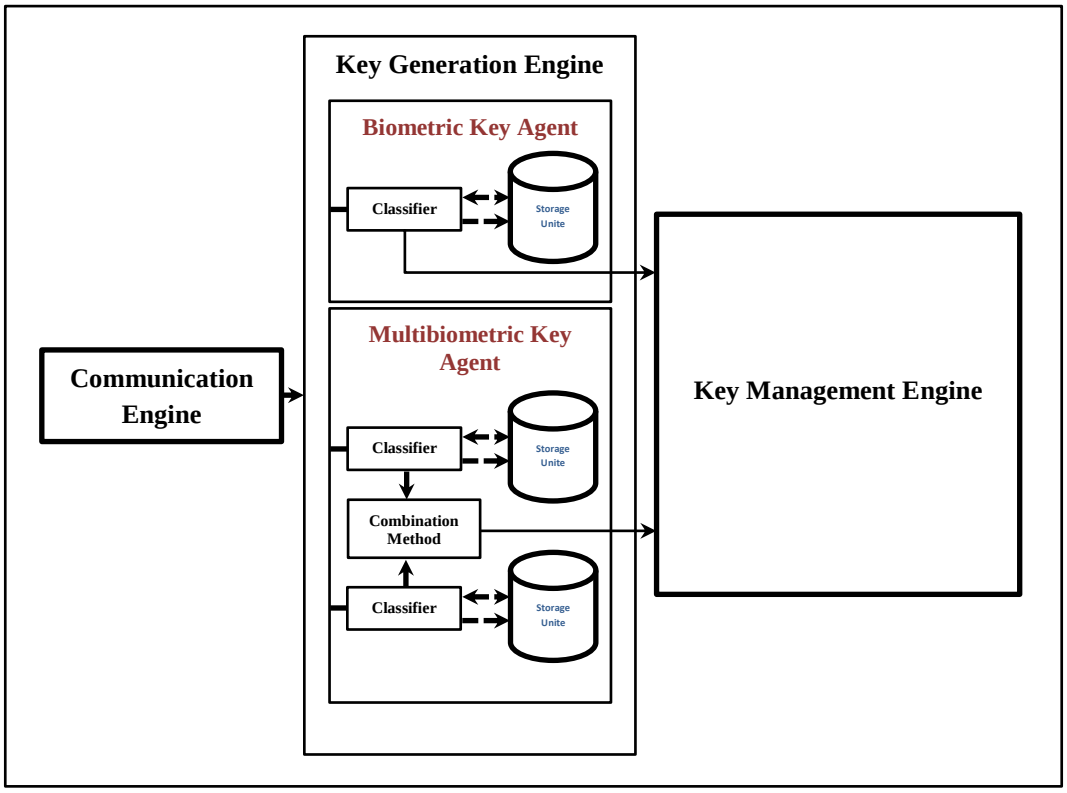


Figure 6. 4 Key Generation Engine

In a **Biometric Key Agent**, the reference biometric features would be inputted into an appropriate classification technique/algorithm to identify a random cryptographic key on training. Thereafter, only the classification parameters on training are stored as a helper data to assist in generating the desired master key over time if necessary. In addition to this, the identified random key on training would be hashed by using a robust hash function and also stored as helper data for key verification purposes. The helper data can be stored locally at some location or would be stored within the Cloud severs if the encryption is undertaken by the cloud provider. It is worth noting that this research presents the possible solution of locally storing data on personal storage units (e.g. hard drives) for facilitating the cryptographic key generation process with a view to stemming and lowering potential attacks. Of course,

local storage units benefit from logical and physical security countermeasures before being able to access data; therefore, they can be considered more secure than cloud-based storage. On the other hand, the existing cloud paradigm lacks the potential afforded to local storage solutions. On encryption/decryption, the classifier structure would be reconfigured once again to generate the master key by using the classification parameters (i.e. the weights in the context of this research) and the live biometric features. In a **Multibiometric Key Agent**, the same previous procedures would be applied in addition to a fusion method as shown in Figure 6.4.

4. **Key Management Engine**

   The **Key Management Engine** represents the central processing unit of the system architecture where it generally controls the biometric key generation in an effective way. This would in particular be accomplished via liaising between both the **Key Generation Engine** and the **Cryptographic Engine**. To this end, the **Key Management Engine** undertakes a number of crucial procedural tasks to manage and maintain a successful bio-crypto key generation. The fundamental tasks/procedures of this engine are outlined by the following:

   - **Key Verification**: The **Key Management Engine** would perform a key verification procedure in order to verify whether the generated master key is valid or not. This would be applied at the encryption/decryption time when the created master key is delivered by the **Key Generation Engine**. For this achievement, a strong cryptographic hash function would be used in a secure way to hash the identified biometric key on registration – the

hashed key will be stored at some location. Then, the master key that is generated on encryption will also be hashed by the same cryptographic hash function which is applied on training. Accordingly, the master key will be correct if its hash value on validation equals the stored hashed key on enrolment. The hash function of SHA-3 can be used to reinforce the security aspect. The hashed key on training would be either stored at a local storage unite or on the Cloud depending on at which side the encryption will be undertaken.

- **Encryption Assets Production**: The **Key Management Engine** also will undertake another procedural task to produce the encryption assets. These assets would accomplish the requirement of generating diverse keys to underpin the privacy and confidentiality of the stored files. That is, each and every single file within the cloud storage account has to be encrypted/decrypted by its own cryptographic key. Thereby, if one file has been hacked the other one would not be affected as it is protected by a different key. Therefore, there is a need to input random seed values along with the valid/correct master key into a reliable random key generator to establish numerous-file keys. As such, a database will be created to index each and every stored file name within cloud storage alongside its seed value. This would be used as a reference to cipher or decipher that file once again by its file key (i.e. unique attribute for each stored file).

- **Key Revocation**: This procedural task generally refers to the capability to cancel a key in case of compromise and reissue another one. Different final bio-crypto keys would be produced by feeding a random key generator approach with a random seed and the correct master key. Accordingly, the **Key Management Engine** includes two levels of revocations. The former is related with revoking the correct master key in case of compromise. The latter, nevertheless, associates with revoking the final bio-crypto keys generated by the **Key Management Engine**. In the foremost revocation scenario, when the master key is leaked, this means that the **Non-Intrusive Biometric Engine** has been hacked. Accordingly, the master key will be cancelled from the bio-cryptosystem. Then, the classification techniques within the **Key Generation Engine** would regenerate another helper data via identifying a new master key on an additional enrolment. In the latter revocation scenario, if any of the final bio-crypto keys is compromised, this could indicate that a random seed has been guessed/revealed. Therefore, the current random seed will be revoked from the system. Following this, another seed would be simply determined and entered into the random key generator along with the master key to produce a new final secret key to tackle this attack. For both revocation scenarios, each file should be encrypted once again by using its own new final secret key. With a view to managing and maintaining secure encryption framework, the master key would be updated on a regular basis.

Using the master key for so long time such as years will certainly present the potential of hacking the biometric data. As such, the **Key Generation Engine** would re-train the classification approaches to recognize different master key and to establish a new helper data on an annually basis or every six months. As explained earlier, in case of keys compromise, the **Key Management Engine** will liaise with the **Key Generation Engine** and the **Cryptographic Engine** to revoke the existing keys and reissue new ones.

5. **Cryptographic Engine**

The **Cryptographic Engine** is responsible for ciphering or deciphering the cloud storage data. In order to achieve this, a number of algorithms can be applied for a robust encryption/decryption goal. The cryptographic algorithms of course can be classified into symmetric and asymmetric cryptographic algorithms, and both of them can achieve secure solutions reliant upon the required application. Amongst cryptographic algorithmic, AES is a reliable symmetric encryption technique; especially AES-256 which has not been broken yet. On the other hand, RSA can be considered a powerful asymmetric encryption algorithm for network security. Whilst symmetric mode is fast and convenient for protecting data at one party, asymmetric mode can be considered slow, but it eradicates the negativity of the latter in terms of sharing the secret key amongst parties. Hence, the asymmetric mode does not clearly represent the core framework in the context of this research. Symmetric encryption then will be employed to resolve the privacy issues within the cloud-based storage. Therefore, the cryptographic engine will receive the final

secret key from the **Key Management Engine**. Accordingly, each file-key is used to seamlessly encrypt each file within the cloud storage by a sophisticated cryptographic algorithm, such as AES-256. At end, the cloud storage provider will be responsible to store the encrypted files.

6.5 Operational Considerations

The core innovative framework introduces a transparent and continuous encryption approach to solve the privacy and confidentiality issues of the stored data within storage service. The design of such architectural system fulfills the fundamental requirements; therefore, it has the foundation for enabling secure and usable encryption in a seamless fashion. Nevertheless, whether the encryption is undertaken at the cloud or at the client side, the presented framework would need further considerations to cope with specific issues that must be put in place for effectively reinforcing the system operation in reality. These considerations are explained and discussed by the following:

- **Privacy and Confidentiality**

  Given the potential of occurring privacy and confidentiality issues regarding the use of biometrical signals, further secure countermeasures are a key factor to be consider. The application of biometric techniques must be achieved in a secure way under which the attempts of attacking the biometric data can be combated. The innovative transparent bio-cryptosystem must provide strategic mechanisms in terms of storing, using, and transmitting the biometric features against potential attacks. This will ensure that the biometrical signals are delivered to the authorized entities only. With regards

to the biometric storage, the biometric samples/features will never be stored anywhere within the core system framework. However, helper/public data is stored at some location to facilitate in creating the bio-crypto key (i.e. neural network weights are not sensitive). This data would be updated on a regular basis to avoid potential attacks.

On the other hand, with the General Data Protection Regulation (GDPR) having gone in effect, cloud providers need to introduce a number of mechanisms aiming at strengthening the privacy of personal data (i.e. collective biometric data) while undertaking the biometric encryption. On the whole, GDPR affords users the control over their data and unifies personal data protection legislation across all Europe no matter where this data is processed (Albrecht, 2016, Tankard, 2016). On that basis, cloud provider has to be committed in contracts, for example, to comply with the GDPR in terms of processing the collected biometric features to undertake Cloud-enabled biometric encryption. A number of standard security protocols would be also set in place with a view to securing the transmission of the biometric signals through the Internet. Accordingly, biometrical signals will be transmitted to the cloud provider by utilizing the Secure Socket Layer (SSL) in a secure fashion (William, 1999).

- **Scalability and Elapsed Time**

  On Cloud-enabled encryption, there are particular issues need to be taken into consideration operationally. That is, the elapsed time of generating a valid bio-crypto key by a legitimate biometric sample (especially the lag of the network communication among the communicative parties) and the storage of helper data have to be considered in a sensible way. The last aspect can

be tackled via the cloud provider where additional storage capacity would be easily scaled/risen up (Han et al., 2012). On the other hand, incorporating additional time lags in a networked solution over a device-centric paradigm may diminish the level of adaptability. Therefore, an attention has to be given with a view to ensuring that the network delay does not have a critical influence upon the entire system operation. The elapsed time is not expected though to be a problematic issue, since the principle of network-based encryption already exists for devices in different network domains (William, 1999). Also, a considerable number of active users are increasingly subscribing on remote services. More importantly, as the cryptography will be undertaken in a non-intrusive and continuous fashion throughout the use of the storage service, the user would not be left waiting.

- Trust

  With the issue of having to trust a third-party on an encryption being enabled at the Cloud, the transmission of sensitive biometric features over the Internet could be considered unreliable procedure. At present, the need for users and organizations to trust a third-party provider do not represent an ideal expectation, and might become a security concern afterwards. On the other hand, clients and organizations already trust service providers with credentials of different security purposes such as verification and encryption. In addition, the user data currently is transmitted to the cloud provider via standard Internet security approaches (e.g. Transport Layer Security (TLS)) (Galibus et al., 2016). Whilst the existing cryptographic solutions protect the stored data, they can defeat to do so against the weak passwords. Access to the data itself is secured via verification passwords approach (i.e. username

and password) (Azam and Johnsson, 2011). With the correct credentials, neither of the security countermeasures managing data at rest or in transit provides any protection – as the system considers the legitimate customer is accessing the data. Accordingly, the novel bio-cryptosystem would offer an effective solution to prevent such attacks without any inconveniences. Despite this, a strong countermeasures should be set in place prior to accessing the service via intrusion detection, for example. Cloud providers at present seek at identifying when an imposter tries to break an account. Thereby, regular levels of detection ensure that cloud providers can prevent intruders who bygone break the network's initial defences (Dobran, 2019). As a result, it can be argued that subscribers and organizations would not be taking any risks within the presented solution – it would be just more of the same type of service as always. Strict statements of service level agreement in addition to continuously monitoring the granted access to the cloud storage would provide a baseline level of trust.

- Helper Data Maintenance

  With cryptography-based biometric being undertaken in a transparent and continuous manner, the biometric registration and renewal stages should be taken into consideration. In view of collecting the biometric samples in a spontaneous manner, the application of transparent biometric techniques can result in a higher degree of feature vector variability. This would significantly impact the behavioural biometric modalities that have time-variant features (e.g. keystroke dynamics) (Clarke, 2011). As such, potential attacks can be overcome by renewing the helper data. It is worth noting that the enrolment maintenance consideration is more related with the processing infrastructure

(i.e. helper data construction is a far more relevant with memory processing than encryption). However, understanding the implication of doing so on a regular basis (i.e. weekly, monthly, or annually) would support the implementation of the transparent bio-cryptosystem in an effective way. For instance, dynamic profile update can be performed to produce helper data via the most biometric samples of the last x sessions as they have been successfully verified in generating the wanted bio-crypto key via the key verification procedure. As such, the entire correct verified samples will be included within the new enrolment phase at the end of the identified period; however, those rejected samples will be used to construct imposter data with a view to re-training the applied classifiers. In this context, clients will have true and robust interactive profile including their up to date biometric features. Another example is that a profile update can be adopted reliant upon biometrics performance. The profile update will retrospectively consider the samples of each biometric approach, if the decisions apart of the fusion method will have been positive with high confidence. On the other hand, the profile update could undertake profile refinement and classification re-training instead, if the decisions will have been negative. A biometric technique of a low accuracy can be also taken into account for a profile update. That is, a biometric modality that could not generate a key for some users can be exploited for profiling update. Whilst this poor biometric approach was fused with other effective biometric approaches, it does not surely mean that the bio-crypto key is generated by the genuine user. Consequently, that biometric technique could be suspended and/or trigger the multibiometric approach for

another training session until gathering the more valid samples to reach a better performance.

- **Multiple Local Storages**

  On local encryption being undertaken at the user side, the helper data storage at different locations might lead to potential vulnerabilities. Currently, this might be perceived as an inappropriate perspective by individuals and enterprises and may become a security issue in the meantime. On the other hand, users and companies are yet storing very important data on different devices ranging from desktop computers, laptops, mobile phones to smart cards. In this context, the service subscribers can understand how sensitive the stored data is and accordingly they would look after the data protection (Chin et al., 2012). The local storage solution exploits logical and physical security controls. As such, the attacks upon particular organizations are clearly not easy. The stored information that could be locally stored in a computer on an office within entire building is protected by secure doors (physical countermeasure) and firewalls (logical countermeasure). Consequently, there are reliable security controls by which threat vectors will be diminished.

  From another perspective, the stored classification parameters in addition to the encryption assets (i.e. the files' names and their random seeds) beyond the biometric features or the master key would be arguably insensitive. The identified biometric key on enrolment/training also will be hashed via strong cryptographic hash function using SHA-3 (Preneel, 2010) to overcome potential attacks. In order to cope with the vulnerabilities that are associated with the biometric features compromise, the helper data construction

alongside the encryption asserts would be updated on a regular basis (i.e. every six months or annually).

- Cost

  The transparent encryption framework presented within this research would improve the cloud-based storage without changing the existing model – enabling more secure and usable cryptography. In return of such an intelligent technology, the provider can sell the cryptographic service to the beneficiaries with a sensible cost instead of the zero-cost strategy which is wrongly perceived to be the cost of many secret-knowledge approaches (Clarke, 2011). As such, the subscriber who transparently takes advantage of the encryption service could be charged a reasonable price per perhaps specific sessions or on a regular basis (e.g. weekly or monthly). On the other hand, the seamless cryptographic service has to be more than viable; thus, the paying conception would be perceived or regarded by subscribers. Incurring cost on encryption is not odd consideration as the cloud providers are currently charging subscribers a fee in exchange for their online storage services. In this context, the cost deems an instrumental factor for adopting the presented service within cloud-based storage where the transparent encryption will prevent the user from having to remember, recall or present difficult and/or tedious cryptographic credentials. These can be secret keys and/or biometric modalities. It also introduces an additional robust user-oriented level of protection to boost the privacy and the confidentiality aspects.

6.6 Discussion

The innovative bio-cryptosystem presents a more secure and usable encryption framework based upon transparent multi-biometric. Accordingly, the biometric samples in the context of this work need to be captured in an entirely non-intrusive manner without having to explicitly interact with the bio-cryptosystem. In particular, whilst the facial and the keystrokes samples will be transparently collected without the constrained interaction of the user, the insight of employing fingerprint towards unconstrained verification (see section 2.6.1.1) will undoubtedly offer mature transparent fingerprint sampling to support the operation of such a novel encryption in reality. On the other hand, incorporating additional transparent biometric modalities into the innovative bio-cryptosystem can possibly boost the reliability of generating a bio-crypto key at all times. For example, voice and gait biometric samples can be collected in a fully non-intrusive fashion while walk or having a call. Although their samples might not be acquired on a timely basis in comparison with the fingerprint, face, and keystrokes samples, they can be considered auxiliary means for eliminating the weaknesses of other biometric techniques sometimes. Therefore, the accuracy of generating the bio-crypto key would be escalated – whereas single biometric approach can impact the effectiveness in terms of high error rate and lack of universality. On that basis, a usable and convenient experience would have been afforded for the cloud storage subscribers by applying multiple transparent biometric modalities, where they no longer need to provide their credentials (e.g. biometrics and secret keys) to encrypt/decrypt the stored data. The multi-biometric approach would also reduce the user inconveniences caused by generating the incorrect key. That is, with numerous samples being collected

transparently form various biometric modalities, the opportunity of creating the correct valid key will be offered as well as this aspect would allow the possibility of trading-off the FRR against the FAR. For further improvement, a number of dynamic profile update procedures would be performed (as explained in section 6.5) to minimize the degree of feature vector variability in particular to those biometric approaches which are behavioural in nature and accordingly have time-variant features, such as keystroke dynamics. As a consequence, the key generation performance will be improved.

With a view to eliminating the potential attacks on biometrics, the biometric features are never stored anywhere at all. Public data (disadvantageous to adversaries) derived from the biometrical signals, however, are stored either at a remote location (i.e. the Cloud) or at some local locations to facilitate in generating the desired bio-crypto key on time. The encryption assets (the files' names and the random seed values for each file) which are arguably insensitive are also stored for producing numerous-file keys to encrypt/decrypt each file within the storage. As a result, the helper data storage in addition to the encryption assets ensure good secure management where an attacker still need the fresh features or the master key to hack the system. Additionally, the identified key on training would be hashed via SHA-3 and stored for key verification purpose afterwards. For revocation, two levels for cancelling leaked keys including the master key and the final-file secret key are presented. In case of compromise, the former will be revoked, and a new master key will be identified via re-training the classification approach. The latter, however, will be cancelled and a fresh final-file secret key will be generated by determining another random seed. With the purpose of resisting brute force attacks, the biometric

entropy has been strengthened by incorporating considerable features from a variety of biometric approaches; therefore, there would be a problematic issue to guess them by forgers. The experiments of Chapter 6 demonstrate a very effective entropy by biometric fusion where the effective bit-length of the multibiometric feature vector reaches up to 6567438592.

Regarding the issues of having to trust a third-parity, users and organisation currently trust service providers with credentials for security aims, and in particular, cloud user data are already transmitted to the cloud provider and vice versa. Whilst current security framework within cloud storage service can be defeated against poor passwords, the innovative cryptographic solution escalates the security and usability of the protection framework within cloud storage. Therefore, clients and organizations arguably wound not be taking any risks by using the proposed transparent encryption – it will be merely more of the same type of service as always. What is more, a reliable standard security protocol can be used in order to protect the transmission of the biometrical signals through the Internet towards the trusted third party.

From the spoofing attack perspective, the forgers will have difficulties in spoofing the transparent biometric samples – hacking the samples which are acquired in a spontaneous fashion is a tricky attempt to take. In addition to this, the transparent bio-cryptosystem architecture implements a multibiometric topology in order to combat the biometric spoofing attack – ruling out the reliance upon single biometric approach that can be considered as an unsatisfactory in terms of low individuality and high forgery attempts. Thereby, the imposter will have no clear capacity to forge all numerous biometric samples collected from different modalities. Furthermore,

updating the user profile as discussed earlier would establish renewed helper data from time to time with a view of representing the true interaction of the valid user with the cloud service and concurrently mitigating the invalid access caused by forging the biometric samples. This will decrease the revocation of the master key as the public data will be updated on a monthly or an annually basis – thus strengthening the security aspect.

For cyberattacks resistance, robust intrusion detection techniques built upon advanced AI developments would be given in place in order to overcome the potential malware attacks which have been hijacking the legitimate biometric samples through the scam of a malicious but masqueraded software on the user's computer. More importantly, genuine users who are quite familiar with technology may be offered cybersecurity education for awareness and compliance in order to have digital secure behaviour. Hence, the cloud-based providers could be in charge of educating their subscribers for having digital secure awareness and compliance against the potential malware attacks.

From the user privacy perspective, using such a transparent bio-cryptographic technique will necessitate the real-time capture of numerous biometric signals across different transparent biometric modalities when the user would continuously interact with his device. Whilst the non-intrusive bio-cryptosystem would reinforce security and usability, it is also recognized that the system will have privacy constraints. This is related to that fact that some users might not be willing their biometric data to be collected by the proposed system. As such, users need to be comfortable with the context of capturing and processing their biometric information if they are seeking for a secure and usable cryptographic framework. However, the

manner of undertaking the transparent bio-cryptographic technique at the user side or at the cloud side has been considered. Thus, the flexibility of undertaking the transparent encryption-based biometric locally at the user side might meet with the perception of those users who have privacy issues in terms of capturing their transparent biometric samples on a regular basis. Despite this, having gone the General Data Protection Regulation (GDPR) in effect (Voigt and Von dem Bussche, 2017), the cloud-based providers must also provide a number of strategies in order to improve and maintain the security and the privacy of personal data which is the collected biometric data in the context of this work. Therefore, the cloud-based providers must comply with the laws of the GDPR with regards to processing the collected biometric features to undertake a bio-cryptographic framework at the cloud side.

From another point of view, the elapsed time of establishing a valid bio-crypto key by a valid biometric sample – in particular the lag of the communicative parties over the network in addition to the scalability helper data storage could impact the acceptability and adaptability of the technology. The former aspect does not appear to be a problematic issue. As the cryptography will be undertaken in a non-intrusive and continuous fashion throughout the use of a service or device, the user would not be left waiting. The latter aspect can be handled by the cloud provider in scaling up further storage capacity.

6.7 Conclusion

The innovative system architecture has offered a non-intrusive and continuous cryptographic framework based upon multiple transparent biometric techniques that

enable a high level of protection and convenience. Despite the fact that the system architecture is not a risk-proof framework, a number of operational aspects have been addressed and critically resolved to be taken into consideration when developing and operating such a bio-cryptosystem in reality. Implicatively, the more the re-enrolment sessions are undertaken on a regular basis, the highly likely the system security and accuracy would be escalated. The profile update techniques discussed in section 6.5 (i.e. helper data maintenance) would also maximise the reliability of generating a timely constant biometric key in a smoothly manner. These techniques would construct new helper data periodically for reflecting the true interaction of the authentic user with the storage service and at the same time would alleviate the illegitimacy attempts in spoofing the biometric samples. From another perspective, whenever the seamless encryption service within the cloud storage technology has been more than viable, the paying cost conception would be perceived or considered by subscribers – resulting in accepting and adopting the innovative framework. This can be evidenced by tracing the sale figures over time.

Chapter Seven: Conclusions and Future Work

7.1 Introduction

The study effectively demonstrates through experimentation a transparent multimodal bio-cryptographic approach for reinforcing the current security framework of cloud-based storage with a high level of convenience. The application of transparent biometrics within bio-cryptography enables more usable and secure encryption. Transparent biometrics eliminates the need of having to remember or present difficult and tedious cryptographic credentials (i.e. secret keys and biometrics).

The implications of this study are presented by highlighting the key contributions and achievements along with the limitations and obstacles encountered during the research; followed by outlining the potential areas that can be investigated in future work.

7.2 Contributions and Achievements

The research overall has accomplished all the objectives which are originally set out in Chapter 1. The core contribution of this work concentrates upon undertaking a series of experimental studies to investigate the concept of the innovative bio-cryptographic approach leading ultimately to the development of transparent bio-cryptosystem architecture. The research establishes the following key contributions and achievements:

- Establishing a comprehensive understanding upon the topics of transparent biometrics and bio-cryptography and in particular contextualizing a number of

transparent biometric approaches with a view to employing the appropriate ones within bio-cryptography for improving the cloud storage technology.

- Critically analysing the prior research of biometric cryptography with regards to existing approaches, strategic schemas, issues, and available solutions.
- Designing and conducting a baseline set of experiments to investigate how reliable the developed bio-cryptographic approach at creating a constant and non-intrusive bio-crypto key from the selective transparent biometric techniques (i.e. fingerprint, face, and keystroke dynamics) on a timely basis.
- Modelling and performing a series of experiments to investigate the influential factors upon the neural network classifier with the aim of improving the key generation accuracy.
- Developing and implementing a number of experiments to explore the more effective biometric features in generating a bio-crypto key of 256-bit length without undermining the entropy factor and accordingly investigating the correlation between the key length (e.g. 128-bit, 256-bit, 512-bit, … etc.) and the accuracy of reproducing the desired key.
- Undertaking and carrying out a set of experiments to explore the potential of improving the key generation accuracy by combining the feature vectors of the applied biometric modalities.
- Designing and conducting a number of experiments to investigate the likelihood of elevating the bio-crypto key generation performance by integrating the matching scores from each classifier being utilized within the individual biometric approach.
- Proposing an innovative transparent bio-cryptosystem architecture based on multibiometric aiming to offer a convenient user-oriented cryptographic

framework to additionally secure the data privacy and confidentiality within cloud storage.

Several papers related to the research have been presented and published in refereed journals and conferences (provided in Appendix A). As a result, the research is deemed having made positive contributions to the domain of cloud storage security and specifically to field of biometric cryptography.

7.3 Limitations of Research

The aim and the objectives of this research have been fulfilled. However, a number of issues associated in particular with the experimentation of this study have been identified under which limitations may have imposed upon the empirical findings in one way or another. These limitations are illustrated by the following:

- There was a limitation existed in experimenting the cryptographic key generation from the fingerprint modality. The limited number of the fingerprint biometric samples prevented a more thorough evaluation of the innovative bio-cryptographic approach. In the experiments of uni-biometric and multi-biometric, there were only 8 samples in total from each respondent. This might have restricted the overall outcomes within the scope topic. Although there was a limitation regarding experimenting few fingerprint biometric data, numerous samples would be spontaneously collected in reality leading to the opportunity of enhancing the key generation effectiveness.
- The selective biometric modalities (fingerprint, face and keystroke dynamics) were combined into the multibiometric secret key generation. Whilst the fingerprint and face datasets were collected from the same participants, the

keystrokes dataset was captured from different volunteers. As a result, the collective samples across the applied biometric databases was not completely from the same user; thus, the data does not present a true reflection of a real user. On the other hand, from a statistical point of view, the incorporative biometric data can be considered valid with a view to validating the proof-of-concept.

- In particular, there was a limitation in conducting the multibiometric approach. In order to set up the experiments, a number of facial and keystroke samples were excluded (i.e. around 33 face samples from some users and 10 keystroke samples from all users) aiming at totalizing them into 50 samples for each modality. At the same time, the 8 fingerprint samples were duplicated into 50 samples to meet the other biometric approaches. However, had the multibiometric approach performed by using the lowest common number of samples only, there would be an unreliable insight about the multibiometric performance. Therefore, this approach is applied to reflect the key generation effectiveness by employing significant biometric samples.

7.4 Future Work

The contribution of this research has enhanced the security and usability issues of cloud storage technology. On the other hand, further research suggestions related to the current scope of the study can be taken into consideration for future work. These suggestions are shown by the following:

- Given a limited number of samples across the selective biometric modalities in general and small number of fingerprint samples in particular, further

research can be undertaken by collecting a considerable number of fingerprint, face, and keystrokes samples in a non-intrusive manner to investigate the transparent key generation from the innovative bio-cryptographic approach within a realistic environment.

- Additional work can be also performed via capturing enormous data of different spectrum of transparent physiological and behavioural biometric modalities in reality to explore the effectiveness of generating the bio-crypto key from the proposed bio-cryptosystem.
- Using alternative pattern classification algorithms within the proposed approach instead of the neural network technique to determine the accuracy of generating the bio-crypto key by such algorithms.

References

ABAZA, A., ROSS, A., HEBERT, C., HARRISON, M. A. F. & NIXON, M. S. 2013. A survey on ear biometrics. ACM computing surveys (CSUR), 45, 22.

ABUGUBA, S., MILOSAVLJEVIC, M. M. & MACEK, N. 2015. An Efficient Approach to Generating Cryptographic Keys from Face and Iris Biometrics Fused at the Feature Level. International Journal of Computer Science and Network Security (IJCSNS), 15, 6.

ADLER, A., YOUMARAN, R. & LOYKA, S. Towards a measure of biometric information. 2006 Canadian Conference on Electrical and Computer Engineering, 2006. IEEE, 210-213.

AK, J., P, F. & AA, R. 2007. Handbook of Biometrics, Springer Science & Business Media.

ALBRECHT, J. P. 2016. How the GDPR will change the world. Eur. Data Prot. L. Rev., 2, 287.

ALVES, D., CRUZ, G. & VINHAL, C. Authentication system using behavioral biometrics through keystroke dynamics. 2014 IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM), 2014. IEEE, 181-184.

ASHBOURN, J. 2004. Practical Biometrics. From Aspiration to Implementation, Berlin et al: Springer.

ATAH, A. J. 2011. Strategies for template-free direct biometric encryption using voice based features. The University of Kent.

ATAH, J. A. & HOWELLS, G. Key generation in a voice based template free biometric security system. European Workshop on Biometrics and Identity Management, 2009. Springer, 170-177.

AZAM, A. & JOHNSSON, M. 2011. Mobile One Time Passwords and RC4 Encryption for Cloud Computing.

BALAKUMAR, P. & VENKATESAN, R. 2011. Secure Biometric Key Generation Scheme for Cryptography using Combined Biometric Features of Fingerprint and Iris. IJCSI International Journal of Computer Science Issues, 8.

BANERJEE, S. P. & WOODARD, D. L. 2012. Biometric authentication and identification using keystroke dynamics: A survey. Journal of Pattern Recognition Research, 7, 116-139.

BANSAL, R., SEHGAL, P. & BEDI, P. 2011. Minutiae extraction from fingerprint images-a review. arXiv preprint arXiv:1201.1422.

BEHL, A. & BEHL, K. An analysis of cloud computing security issues. Information and Communication Technologies (WICT), 2012 World Congress on, 2012. IEEE, 109-114.

BIOMETRICSINSTITUTELIMITED 2013. Biometrics Institute Industry Survey 2013 – Executive Summary.

BISCHOFF, P. 2018. The best apps to encrypt your files before uploading to the cloud [Online]. Available: https://www.comparitech.com [Accessed January 28 2019].

BODO, A. 1994. Method for producing a digital signature with aid of a biometric feature. German patent DE, 42, 908.

BOULGOURIS, N. V., PLATANIOTIS, K. N. & MICHELI-TZANAKOU, E. 2009. Biometrics: theory, methods, and applications, John Wiley & Sons.

BUTLER, B. 2013. Gartner: Top 10 cloud storage providers [Online]. NETWORKWORLD. Available: http://www.networkworld.com/ [Accessed May 20, 2015].

CAMPBELL, J. P. 1997. Speaker recognition: a tutorial. Proceedings of the IEEE, 85, 1437-1462.

CAVOUKIAN, A. & STOIANOV, A. 2007. Biometric encryption: A positive-sum technology that achieves strong authentication, security and privacy, Information and Privacy Commissioner, Ontario.

CECCARELLI, A., MONTECCHI, L., BRANCATI, F., LOLLINI, P., MARGUGLIO, A. & BONDAVALLI, A. 2015. Continuous and transparent user identity verification for secure internet services. IEEE Transactions on Dependable and Secure Computing, 12, 270-283.

CHANG, E.-C. & ROY, S. Robust extraction of secret bits from minutiae. International Conference on Biometrics, 2007. Springer, 750-759.

CHANG, T.-Y. 2012. Dynamically generate a long-lived private key based on password keystroke features and neural network. Information Sciences, 211, 36-47.

CHAWLA, N. V., BOWYER, K. W., HALL, L. O. & KEGELMEYER, W. P. 2002. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

CHIN, E., FELT, A. P., SEKAR, V. & WAGNER, D. Measuring user confidence in smartphone security and privacy. Proceedings of the eighth symposium on usable privacy and security, 2012. ACM, 1.

CHIN, Y. J., ONG, T. S., TEOH, A. B. J. & GOH, K. 2014. Integrated biometrics template protection technique based on fingerprint and palmprint feature-level fusion. Information fusion, 18, 161-174.

CHUNG, Y., MOON, D., LEE, S., JUNG, S., KIM, T. & AHN, D. Automatic alignment of fingerprint features for fuzzy fingerprint vault. International Conference on Information Security and Cryptology, 2005. Springer, 358-369.

CIF. 2019. Cloud: Driving Business Transformation [Online]. CloudIndustryForum. Available: https://www.cloudindustryforum.org [Accessed].

CIMATO, S., GAMASSI, M., PIURI, V., SASSI, R. & SCOTTI, F. A multi-biometric verification system for the privacy protection of iris templates. Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08, 2009. Springer, 227-234.

CLARKE, N. 2011. Transparent user authentication: biometrics, RFID and behavioural profiling, Springer Science & Business Media.

CLARKE, N. & FURNELL, S. 2005. Biometrics–The promise versus the practice. Computer Fraud & Security, 2005, 12-16.

CLARKE, N., KARATZOUNI, S. & FURNELL, S. Transparent facial recognition for mobile devices. Proceedings of the 7th Security Conference, Las Vegas, 2008. Citeseer.

COLUMBUS, L. 2016. Roundup of cloud computing forecasts and market estimates, 2016. Forbes.

CONGRESS, N. R. T. T. U. S. 2012. Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability.

CRISP, S. 2013. Camera sensor size: Why does it matter and exactly how big are they? [Online]. NEW ATLAS. Available: http://newatlas.com/camera-sensor-size-guide/26684/ [Accessed October 12, 2015].

DANIELYAN, E. 2004. The lures of biometrics. The Internet Protocol Journal, 7, 15-35.

DOBRAN, B. 2019. Cloud Storage Security: How Secure is Your Data in The Cloud? [Online]. PhoenixNAP. Available: https://phoenixnap.com/blog/cloud-storage-security [Accessed 23 May 2019 2019].

DODIS, Y., REYZIN, L. & SMITH, A. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. International Conference on the Theory and Applications of Cryptographic Techniques, 2004. Springer, 523-540.

DRAGO, I., MELLIA, M., M MUNAFO, M., SPEROTTO, A., SADRE, R. & PRAS, A. Inside dropbox: understanding personal cloud storage services. Proceedings of the 2012 ACM conference on Internet measurement conference, 2012. ACM, 481-494.

FENG, G.-C. & YUEN, P. C. 1998. Variance projection function and its application to eye detection for human face recognition. Pattern Recognition Letters, 19, 899-906.

FENG, Y. C. & YUEN, P. C. 2012. Binary discriminant analysis for generating binary face template. IEEE Transactions on Information Forensics and Security, 7, 613-624.

GALIBUS, T., KRASNOPROSHIN, VIKTOR V., ALBUQUERQUE, ROBSON DE O. & FREITAS, EDISON PIGNATON D. 2016. Elements of Cloud Storage Security Concepts, Designs and Optimized Practices. Available: https://www.researchgate.net/.

GANNES, L. 2013. With 120M Users, Google Drive Gets Tighter Integration With Gmail [Online]. All Thingsd. Available: http://allthingsd.com/ [Accessed May 12, 2015].

GARRIS, M. D. & MCCABE, R. M. 2000. NIST special database 27: Fingerprint minutiae from latent and matching tenprint images. National Institute of Standards and Technology, Technical Report NISTIR, 6534.

GILDRED, J. 2018. Dropbox vs Google Drive: the Battle of the Titans [Online]. Cloudwards. Available: https://www.cloudwards.net/dropbox-vs-google-drive/ [Accessed 15th September 2018 2018].

GIOT, R., EL-ABED, M. & ROSENBERGER, C. 2009. GREYC Keystroke: a Benchmark for Keystroke Dynamics Biometric Systems. IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS 2009). Washington, District of Columbia, USA: IEEE Computer Society.

GOLIC, J. D. & BALTATU, M. 2008. Entropy analysis and new constructions of biometric key generation systems. IEEE Transactions on Information Theory, 54, 2026-2040.

GONZALEZ, R. C. & WOODS, R. E. 2008. Digital image processing. Nueva Jersey.

GRIFFITH, E. 2014. Who's winning the consumer cloud storage wars? [Online]. Fortune. Available: http://fortune.com/ [Accessed September 9, 2015].

HADID, A., EVANS, N., MARCEL, S. & FIERREZ, J. 2015. Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. IEEE Signal Processing Magazine, 32, 20-30.

HAGAN, M. T., DEMUTH, H. B., BEALE, M. H. & DE JESÚS, O. 1996. Neural network design, Pws Pub. Boston.

HAN, R., GUO, L., GHANEM, M. M. & GUO, Y. Lightweight resource scaling for cloud applications. Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012), 2012. IEEE Computer Society, 644-651.

HAO, F., ANDERSON, R. & DAUGMAN, J. 2006. Combining crypto with biometrics effectively. IEEE transactions on computers, 55, 1081-1088.

HATTERSLEY, L. 2016. How to use Touch ID: best way to set up finger scanning on your iPhone so it always works [Online]. Macworld. Available: http://www.macworld.co.uk/ [Accessed May 30, 2016].

HE, H. & GARCIA, E. A. 2008. Learning from imbalanced data. IEEE Transactions on Knowledge & Data Engineering, 1263-1284.

HOLDER, E. H., ROBINSON, L. O. & LAUB, J. H. 2011. The fingerprint sourcebook, US Department. of Justice, Office of Justice Programs, National Institute of Justice.

HOQUE, S., FAIRHURST, M. & HOWELLS, G. Evaluating biometric encryption key generation using handwritten signatures. Bio-inspired Learning and Intelligent Systems for Security, 2008. BLISS'08. ECSIS Symposium on, 2008. IEEE, 17-22.

HUANG, J. & WECHSLER, H. 1999. Eye detection using optimal wavelet packets and radial basis functions (rbfs). International Journal of Pattern Recognition and Artificial Intelligence, 13, 1009-1025.

INTHAVISAS, K. & SUNGPRASERT, N. 2013. Speech Cryptographic Key Generation. International Journal of Computer Science and Electronics Engineering (IJCSEE), 1, 326-329.

JAGADEESAN, A. & DURAISWAMY, K. 2010. Secured cryptographic key generation from multimodal biometrics: feature level fusion of fingerprint and iris. arXiv preprint arXiv:1003.1458.

JAIN, A., FLYNN, P. & ROSS, A. A. 2007. Handbook of biometrics, Springer Science & Business Media.

JAIN, A. K., NANDAKUMAR, K. & NAGAR, A. 2008. Biometric template security. EURASIP Journal on Advances in Signal Processing, 2008, 113.

JAIN, A. K., ROSS, A. & PRABHAKAR, S. 2004. An introduction to biometric recognition. IEEE Transactions on circuits and systems for video technology, 14, 4-20.

JANBANDHU, P. K. & SIYAL, M. Y. 2001. Novel biometric digital signatures for Internet-based applications. Information Management & Computer Security, 9, 205-212.

JU, J., WU, J., FU, J., LIN, Z. & ZHANG, J. 2011. A survey on cloud storage. Journal of Computers, 6, 1764-1771.

JUELS, A. & SUDAN, M. 2002. A Fuzzy Vault Scheme. IEEe

JUELS, A. & WATTENBERG, M. A fuzzy commitment scheme. Proceedings of the 6th ACM conference on Computer and communications security, 1999. ACM, 28-36.

KALSOOM, S. & ZIAUDDIN, S. 2012. Iris Recognition: Existing Methods and Open Issues. The Fourth International Conferences on Pervasive Patterns and Applications.

KANADE, S., CAMARA, D., KRICHEN, E., PETROVSKA-DELACRÉTAZ, D. & DORIZZI, B. Three factor scheme for biometric-based cryptographic key regeneration using iris. Biometrics Symposium, 2008. BSYM'08, 2008. IEEE, 59-64.

KANADE, S., CAMARA, D., PETROVSKA-DELACRTAZ, D. & DORIZZI, B. 2009a. Application of biometrics to obtain high entropy cryptographic keys. World Acad. Sci. Eng. Tech, 52, 330.

KANADE, S., PETROVSKA-DELACRÉTAZ, D. & DORIZZI, B. Multi-biometrics based cryptographic key regeneration scheme. Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on, 2009b. IEEE, 1-7.

KAROVALIYA, M., KAREDIA, S., OZA, S. & KALBANDE, D. 2015. Enhanced security for ATM machine with OTP and Facial recognition features. Procedia Computer Science, 45, 390-396.

KHALIL-HANI, M., MARSONO, M. N. & BAKHTERI, R. 2013. Biometric encryption based on a fuzzy vault scheme with a fast chaff generation algorithm. Future Generation Computer Systems, 29, 800-810.

KOUNDINYA, P., THERIL, S., FENG, T., PRAKASH, V., BAO, J. & SHI, W. Multi resolution touch panel with built-in fingerprint sensing support. 2014 Design, Automation & Test in Europe Conference & Exhibition (DATE), 2014. IEEE, 1-6.

KOVACH, S. 2014. Nearly 7 Million Dropbox Passwords Have Been Hacked [Online]. Business Insider. Available: http://www.businessinsider.com/ [Accessed May 8, 2015].

KRIG, S. 2014. Image Pre-Processing. Computer Vision Metrics. Springer.

LE, C. & JAIN, R. 2009. A survey of biometrics security systems. EEUU. Washington University in St. Louis.

LI, C., HU, J., PIEPRZYK, J. & SUSILO, W. 2015. A new biocryptosystem-oriented security analysis framework and implementation of multibiometric cryptosystems based on decision level fusion. IEEE transactions on Information Forensics and Security, 10, 1193-1206.

LI, P., YANG, X., QIAO, H., CAO, K., LIU, E. & TIAN, J. 2012. An effective biometric cryptosystem combining fingerprints with error correction codes. Expert Systems with Applications, 39, 6562-6574.

LI, Q., SUTCU, Y. & MEMON, N. Secure sketch for biometric templates. International Conference on the Theory and Application of Cryptology and Information Security, 2006. Springer, 99-113.

LIM, M.-H. & YUEN, P. C. 2016. Entropy measurement for biometric verification systems. IEEE transactions on cybernetics, 46, 1065-1077.

LOUPPE, G. 2014. Understanding random forests: From theory to practice. arXiv preprint arXiv:1407.7502.

LUXAND. 2016. Luxand Face Recognition [Online]. Available: https://www.luxand.com/apps/facerecognition/ [Accessed 26th March 2016].

MALTONI, D., MAIO, D., JAIN, A. & PRABHAKAR, S. 2009. Handbook of fingerprint recognition, Springer Science & Business Media.

MARIÑO, R. Á., ÁLVAREZ, F. H. & ENCINAS, L. H. 2012. A crypto-biometric scheme based on iris-templates with fuzzy extractors. Information Sciences, 195, 91-102.

MARKETSANDMARKETS. 2016. Global Biometrics Technology Market (2016-2023) – Market Forecast by Products, End-User Application and Geography [Online]. Available: http://www.marketsandmarkets.com/ [Accessed May 12, 2018].

MCCUE, J. 2014. Cloud Computing: United States Businesses Will Spend $13 Billion On It [Online]. Forbes. Available: http://www.forbes.com/ [Accessed May 10, 2015].

MELLOR, C. 2016. Public cloud storage spend to double in two years – reliable sources [Online]. The Register. Available: http://www.theregister.co.uk/ [Accessed September 8, 2016].

MERKLE, J., KEVENAAR, T. & KORTE, U. Multi-modal and multi-instance fusion for biometric cryptosystems. Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the, 2012. IEEE, 1-6.

MOHAMMADZADE, H., SAYYAFAN, A. & GHOJOGH, B. 2018. Pixel-level alignment of facial images for high accuracy recognition using ensemble of patches. JOSA A, 35, 1149-1159.

MOHANAVALLI, S. S., ANAND, S. & HARIHARAN, S. 2014. Biometric Key Generation Using Fingerprints. Australian Journal of Basic and Applied Sciences, 8, 516-524.

MONROSE, F. & RUBIN, A. D. 2000. Keystroke dynamics as a biometric for authentication. Future Generation computer systems, 16, 351-359.

MONWAR, M. 2013. A Multimodal Biometric System Based on Rank Level Fusion. University of Calgary.

NANAVATI, S., THIEME, M. & NANAVATI, R. 2002. Biometrics, Identity Verification in a Networked World, Wiley Computer Publishing, 2002.

NANDAKUMAR, K. 2008. Multibiometric systems: Fusion strategies and template security. MICHIGAN STATE UNIV EAST LANSING DEPT OF COMPUTER SCIENCE/ENGINEERING.

NANDAKUMAR, K., JAIN, A. K. & PANKANTI, S. 2007. Fingerprint-based fuzzy vault: Implementation and performance. IEEE transactions on information forensics and security, 2, 744-757.

NEUROTECHNOLOGY. 2016. Fingerprint identifi cation for stand-alone or Web solutions, VeriFinger SDK [Online]. Available: WWW.NEUROTECHNOLOGY.com [Accessed 5th February 2016].

NGUYEN, T. H., WANG, Y., HA, Y. & LI, R. 2013. Improved chaff point generation for vault scheme in bio-cryptosystems. IET biometrics, 2, 48-55.

NT, B. 2014. 5 advantages and disadvantages of Cloud Storage [Online]. BIG DATA. Available: http://bigdata-madesimple.com/ [Accessed May 11, 2015].

OSHIRO, T. M., PEREZ, P. S. & BARANAUSKAS, J. A. How many trees in a random forest? International Workshop on Machine Learning and Data Mining in Pattern Recognition, 2012. Springer, 154-168.

PANDA & RANJAN, D. 2007. Eye detection using wavelets and ANN. National Institute of Technology Rourkela.

PAREKH, M. D. H. & SRIDARAN, R. 2013. An Analysis of Security Challenges in Cloud Computing. IJACSA) International Journal of Advanced Computer Science and Applications, 4.

PENG, K., CHEN, L., RUAN, S. & KUKHAREV, G. 2005. A robust agorithm for eye detection on gray intensity face without spectacles. Journal of Computer Science & Technology, 5.

PFLUG, A. & BUSCH, C. 2012. Ear biometrics: a survey of detection, feature extraction and recognition methods. IET biometrics, 1, 114-129.

PHILLIPSON, C. 2016. CLOUD STORAGE FOR BUSINESS: 37 CLOUD EXPERTS REVEAL THE TOP CLOUD STORAGE MISTAKES THEY SEE COMPANIES MAKE [Online]. All Things Productivity Available: http://www.docurated.com/all-things-productivity [Accessed].

PRENEEL, B. The first 30 years of cryptographic hash functions and the NIST SHA-3 competition. Cryptographers' track at the RSA conference, 2010. Springer, 1-14.

RATHGEB, C. & BUSCH, C. 2012. Multi-biometric template protection: Issues and challenges, INTECH Open Access Publisher.

RATHGEB, C. & UHL, A. 2011. A survey on biometric cryptosystems and cancelable biometrics. EURASIP Journal on Information Security, 2011, 1.

ROBERTS, C. 2007. Biometric attack vectors and defences. Computers & Security, 26, 14-25.

ROSS, A. A., NANDAKUMAR, K. & JAIN, A. 2006. Handbook of multibiometrics, Springer Science & Business Media.

SHANNON, C. E. & WEAVER, W. 1949. The mathematical theory of communication (Urbana, IL. University of illinois Press IL.

SHENG, W., HOWELLS, G., FAIRHURST, M. & DERAVI, F. 2008. Template-free biometric-key generation by means of fuzzy genetic clustering. IEEE Transactions on Information Forensics and Security, 3, 183-191.

SIM, T., ZHANG, S., JANAKIRAMAN, R. & KUMAR, S. 2007. Continuous verification using multimodal biometrics. IEEE transactions on pattern analysis and machine intelligence, 29, 687-700.

SNELICK, R., INDOVINA, M., YEN, J. & MINK, A. Multimodal biometrics: issues in design and testing. Proceedings of the 5th international conference on Multimodal interfaces, 2003. ACM, 68-72.

SOLAYAPPAN, N. & LATIFI, S. A survey of unimodal biometric methods. Proceedings of the 2006 International Conference on Security and Management, 2006. 57-63.

SONG, O. T., JIN, A. T. B. & NGO, D. C. L. 2008. Application-Specific Key Release Scheme from Biometrics. IJ Network Security, 6, 127-133.

SOUTAR, C., ROBERGE, D., STOIANOV, A., GILROY, R. & KUMAR, B. V. Biometric encryption using image processing. Photonics West'98 Electronic Imaging, 1998. International Society for Optics and Photonics, 178-188.

SOUTAR, C., ROBERGE, D., STOIANOV, A., GILROY, R. & KUMAR, B. V. 1999. Biometric Encryption, chapter 22 in ICSA Guide to Cryptography. McGraw-Hill.

SPACEK, L. 2007. Collection of facial images: Faces94. Computer Vision Science and Research Projects, University of Essex, United Kingdom, http://cswww. essex. ac. uk/mv/allfaces/faces94. html.

STANDARDIZATION, I. O. F. 2005. Information Technology, Biometric Data Interchange Formats, ISO/IEC.

STEVENS, M. M. J. 2012. Attacks on hash functions and applications, Mathematical Institute, Faculty of Science, Leiden University.

STOIANOV, A. Security of error correcting code for biometric encryption. Privacy Security and Trust (PST), 2010 Eighth Annual International Conference on, 2010. IEEE, 231-235.

STORMANN, C. 1997. Fraud management tool: evaluation report. Advanced Security for Personal Communications (ASePECT), Deliverable. 13, Doc Ref. AC095/SAG.

SUKARNO, P., BHATTACHARJEE, N. & SRINIVASAN, B. An effective crypto-biometric system for secure email in wireless environment. Proceedings of the 7th International Conference on Advances in Mobile Computing and Multimedia, 2009. ACM, 241-245.

SULLIVAN, D. 2015. Dropbox vs Google Drive [Online]. Cloudwards Available: http://www.cloudwards.net/ [Accessed May 12, 2016].

SULTANA, M., PAUL, P. P. & GAVRILOVA, M. A concept of social behavioral biometrics: Motivation, current developments, and future trends. Cyberworlds (CW), 2014 International Conference on, 2014. IEEE, 271-278.

SUTCU, Y., LI, Q. & MEMON, N. Secure biometric templates from fingerprint-face features. 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007. IEEE, 1-6.

SUTCU, Y., RANE, S., YEDIDIA, J. S., DRAPER, S. C. & VETRO, A. Feature transformation of biometric templates for secure biometric systems based on error correcting codes. Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on, 2008. IEEE, 1-6.

TANKARD, C. 2016. What the GDPR means for businesses. Network Security, 2016, 5-8.

TEOH, A. B. J. & KIM, J. 2007. Secure biometric template protection in fuzzy commitment scheme. IEICE Electronics Express, 4, 724-730.

TIAN, L.-Q., LIN, C. & NI, Y. Evaluation of user behavior trust in cloud computing. 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), 2010. IEEE, V7-567-V7-572.

TOMKO, G. J., SOUTAR, C. & SCHMIDT, G. J. 1996. Fingerprint controlled public key cryptographic system. Google Patents.

TOORANI, M. & BEHESHTI, A. SSMS-A secure SMS messaging protocol for the m-payment systems. 2008 IEEE Symposium on Computers and Communications, 2008. IEEE, 700-705.

TOTH, B. 2005. Biometric liveness detection. Information Security Bulletin, 10, 291-297.

ULUDAG, U., PANKANTI, S., PRABHAKAR, S. & JAIN, A. K. 2004. Biometric cryptosystems: issues and challenges. Proceedings of the IEEE, 92, 948-960.

USMAN, A. K. & SHAH, M. H. 2013. Strengthening e-banking security using keystroke dynamics. Journal of Internet Banking and Commerce, 18.

VOIGT, P. & VON DEM BUSSCHE, A. 2017. The EU General Data Protection Regulation (GDPR). A Practical Guide, 1st Ed., Cham: Springer International Publishing.

VONNIE. 2014. 5 million gmail passwords leaked today Here are 4 actions you need to take [Online]. FixedByVonnie.Com. Available: http://www.fixedbyvonnie.com/ [Accessed May 31, 2015].

WAHDAN, H., WAHDAN, A.-M. & YOUSSIF, A. A. 2013. CRYPTOSYSTEM FROM MULTIPLE BIOMETRIC MODALITIES. European Scientific Journal, 9.

WILLIAM, S. 1999. Cryptography and network security: principles and practice. Prentice-Hall, Inc, 23-50.

WOODWARD, J. D., ORLANS, N. M. & HIGGINS, P. T. 2003. Biometrics:[identity assurance in the information age], McGraw-Hill/Osborne New York.

XIE, X., SUDHAKAR, R. & ZHUANG, H. 1994. On improving eye feature extraction using deformable templates. Pattern Recognition, 27, 791-799.

YAMAZAKI, Y. & KOMATSU, N. 2001. A secure communication system using biometric identity verification. IEICE TRANSACTIONS on Information and Systems, 84, 879-884.

YAMPOLSKIY, R. V. 2008. Behavioral modeling: an overview. American Journal of Applied Sciences, 5, 496-503.

YAN, P. & BOWYER, K. Empirical evaluation of advanced ear biometrics. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops, 2005. IEEE, 41-41.

YEE, L. T. 2011. HARDWARE-BASED BIOMETRIC ENCRYPTION IMPLEMENTATION WITH GAUSS-JORDAN ALGORITHM ACCELERATOR CORE IN FIELD PROGRAMMABLE GATE ARRAYS.

YIN, Y., LIU, L. & SUN, X. SDUMLA-HMT: a multimodal biometric database. Chinese Conference on Biometric Recognition, 2011. Springer, 260-268.

ZIAUDDIN, S. & DAILEY, M. N. 2010. Robust iris verification for key management. Pattern Recognition Letters, 31, 926-935.